

Semantički Web, sustavi za organizaciju znanja i mrežni standardi

AIDA SLAVIĆ*

Sažetak

Internet se ubrzano kreće prema novom razvojnom stupnju (Internet 2) zasnovanom na WWW arhitekturi poznatoj pod nazivom Semantički Web. U ovom će okruženju Web postupno biti popunjen sadržajima s formalnom izraženom semantikom što će omogućiti potpunu automatizaciju u organizaciji i pronalaženju informacija. Kao što je očigledno već iz naziva ova vrsta mreže podrazumijeva formalno iskazivanje veze između dokumenata i njihovog značenja (semantike) kako bi organizacija informacija i informacijske službe mogle u cijelosti funkcionirati putem računalne komunikacije (m2m). Poradi toga Semantički web je neposredno vezan uz razvoj tzv. ontologijskih rječnika. Tradicionalni sustavi za organizaciju znanja, poput klasifikacija i tezaurusa, već se duže vremena koriste u organizaciji i pronalaženju informacija na Internetu posebno u izgradnji predmetnih direktorija. Važnost ovih pomagala posve je izvjesna u kontekstu Semantičkog weba i ovisit će o tome da li se oni mogu koristiti i razmjenjivati putem nekog od standarda za razmjenu kontroliranih rječnika zasnovanih na principu računalne ontologije.

Ključne riječi: semantički Web, ontologija, sustavi za organizaciju znanja, kontroliran rječnici, pristup i razmjena rječnika, mrežni standardi, formati.

1. Uvodno o semantičkom Webu

Govoreći o pronalaženju informacija 1996, Berners-Lee je posebno istaknuo činjenicu kako je vrlo teško pronaći određeni podatak ukoliko on nije izražen u obliku konkretnog semantičkog navoda, tj. u obliku u kojem je znanje prikazano kroz neki jezični navod (Berners-Lee, 1996). U narednim godinama na konfe-

* Aida Slavić, School of Library, Archives and Information Studies, University College London. E-mail: aida@acorweb.net

rencijama WWW 7 (Brisbane, 1997) i WWW 8 (Toronto, 1998) on je izložio ideju globalnog Weba sa samostalnom unutarnjom logikom i službeno najavio sljedeću fazu razvoja nazvanu semantički Web. Time pronalaženje informacija ulazi u novu dimenziju budući da je svrha semantičkog Weba povezati temeljna područja ljudskog znanja i omogućiti njihovu strojnu obradu. Ključni dio čitave ideje sadržan je u semantici predmetnih metapodataka (*subject metadata*) i njihovom prikazu kroz kontekstualiziranu i stroju “razumljivu” terminologiju. Osnovni smisao ideje semantičkog Weba kako ga izražava Berners-Lee, je u tome da će strojno obradivi, “smisleni” metapodaci biti osnova za nadolazeći naraštaj informacijskih službi za pronalaženje informacija koje će i ljudima i računalima olakšati pristup informacijama i komunikaciju. Tako će, primjerice, biti moguće kreirati inteligentne programske agente (*intelligent agents*) koji će čak i vrlo složene zadatke moći izvršavati sasvim samostalno pomoću “m2m” (*machine to machine*) komunikacije (Berners-Lee, Hendlers & Lassila, 2001; Veltman, 2001, 2004).

Dok su danas izvori informacija na Webu uglavnom namijenjeni ljudskom korisniku koji njima upravlja posredstvom jednostavnih veza (linkova), u budućnosti će podatke obrađivati programi kreirani neovisno od podataka. Za te će programe biti neophodni strojno čitljivi navodi o izvorima informacija i njihovim međusobnim odnosima koji će se temeljiti na: postojanju jednog standardiziranog modela, na vezi između pojmova u rječniku i njihovih jedinstvenih definicija i na dostupnosti tih definicija programima. Time se podrazumijeva da će programski agenti upravljati Webom koji će se sastojati od opisa i “ontologija” (uključujući još nesastavljene rječnike), samostalno obrađujući prikupljene podatke i komunicirajući putem djelomičnog razumijevanja (*partial understanding*).

Ideja semantičkog Weba podrazumijeva da će se Internet moći pretraživati ne samo korištenjem riječi, već i kroz upotrebu značenja. To očigledno zahtijeva i semantičku i sintaktičku interoperabilnost predmetnog rječnika, budući da je dobro poznato da su za temeljit i sveobuhvatan opis predmeta potrebni ne samo izolirani pojmovi već i propozicijska logika (Veltman, 2001). U tom kontekstu, postojeće sustave za organizaciju znanja (*knowledge organization systems* skr. KOS) kao što su, na primjer, klasifikacijski sustavi, prepoznalo se kao važne izvore strukturiranih i formaliziranih rječnika koji mogu biti izuzetno korisni u razvoju semantičkog Weba. Neovisno o korištenoj indeksnom pojmu (notacijski simbol ili riječ) sustavi za organizaciju znanja prepoznatljivi su po logičkim procesima koje uključuju, po svojoj strukturi ili prema funkcijama i načinu prikaza znanja. Onima koji rade na primjeni informacijskih sustava od velike su koristi, na primjer, klasifikacijske strukture koje se mogu upotrijebiti pri izradi informacijskih mapa, pojmovnih mapa, kod vizualizacije pristupa predmetu, prikaza interaktivnog pretraživanja, te za pregled distribuiranih iz-

vora informacija¹. Svaka od ovih primjena je usko povezana sa dostupnošću klasifikacijskih podataka u obliku u kojem ih stroj može samostalno obraditi.

Dvije su razvojne tendencije na području sustava za organizaciju znanja, posebice unutar konteksta klasifikacijskih struktura znanja, bitne za razvoj ideje semantičkog Weba i vrlo će vjerojatno utjecati na buduću upotrebu sustava za organizaciju znanja: 1.) standardi i programska rješenja za terminološku razmjenu; 2.) ontologijski rječnici (u obliku ontologija s područja umjetne inteligencije). Uključujući ova dva razvojna područja, proces standardizacije se primarno usredotočuje na tehnološke okvire mreže i na daljnji napredak u razvoju mrežnih jezika za reprezentaciju sadržaja (uz korištenje XML i XML/RDF² sustava kodiranja).

2. Ontologije i jezici za formatiranje ontologijskih rječnika na Webu

Unutar ideje semantičkog Weba velika važnost pridaje se unapređenju strojno obradivih struktura znanja što je izravno vezanu uz područje inženjeringa znanja (*knowledge engineering*) i umjetne inteligencije (*artificial intelligence*). Inženjeri informacijskih sustava koji rade na sastavljanju i usvajanju strojno razumljivih, javno dostupnih pojmovnih rječnika najčešće se služe standardima i formatima za razmjenu podataka koje razvija World Wide Web Consortium (W3C) ili koji druga zajednica mrežnih korisnika (Noy & McGinness, 2001). Od 1980-ih naovamo važnost sustava utemeljenih na znanju (*knowledge based systems* – KBS) često je bila predmet istraživanja informacijskih znanosti i to najčešće u kontekstu automatskog označivanja i pronalaženja informacija (Croft, 1997). Budući je mogućnost dobivanja cjelovitog teksta danas ključna u pronalaženju informacija putem Interneta, specifični pojmovi kao što su “domena znanja” (*knowledge domain*) i “ontologija” sve se češće upotrebljavaju u kontekstu informacijske znanosti (Vickery, 1997). Pojam “ontologija” postupno počinje obuhvaćati čitav niz značenja i uključuje sve od taksonomija, kontroliranih rječnika koji se koriste u metapodacima, popisa proizvoda ili klasifikacija usluga, do rječnika baza podataka i njihovih odnosa. Ontologija u smislu for-

¹ Soergel (1999: 1119) sažima sljedeće uloge knjižnične klasifikacije u mrežnom okruženju: predstavlja semantičke karte; poboljšava komunikaciju i učenje; pruža konceptijsku osnovu za plan istraživanja; pruža klasifikaciju za djelovanje i aktivnosti; poboljšava dostupnost informacija; osigurava konceptijsku osnovu za sustave znanja (*knowledge based systems*); pruža konceptijsku osnovu za definiciju informacijskih elemenata i hijerarhije objekata unutar programskih sustava; služi u izradi interdisciplinarnih, međujezičnih i interkulturalnih karata; služi kao jedno-, dvo- ili višejezična osnova rječnika ili ukupnog znanja u obradi prirodnog jezika.

² RDF omogućava kodiranje, razmjenu i višestruku upotrebu strukturiranih metapodataka, te kombiniranje nekoliko različitih struktura metapodataka unutar jednog jedinog zapisa koja ima mogućnost uspostavljanja veze sa vanjskim izvorom referenci. U identificiranju izvora i uspostavljanju veza među njima RDF se oslanja na URI-a (*Uniform Resource Identifier*) i.e. jedinstveni identifikator izvora podataka.

malne strukture podataka koja se koristi u izgradnji baze znanja je čak i unutar ovog područja relativno nov predmet istraživanja i datira iz 1990-ih (Ding, 2001; Vickery, 1997).

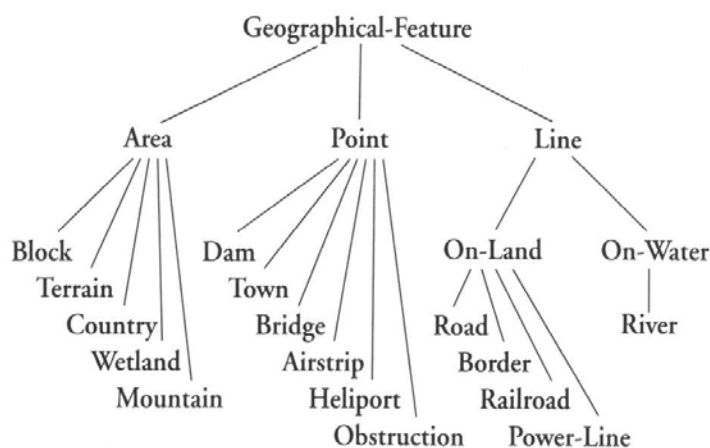
2.1 Ontologija u računarstvu

Ontologija je usko povezana sa sustavima baza znanja (*knowledge based systems – KBS*) tj. ekspertnim sustavima koji su programirani da se “ponašaju inteligentno” i na taj način ili pomognu ljudskim stručnjacima da izvrše svoje zadaće brže i ekonomičnije, ili u potpunosti zamijene ljude u izvođenju opasnih i skupih rutinskih procedura. Takav računalni sustav potrebno je opremiti znanjem iz određenog područja tj. baze znanja (*knowledge base*) i programirati ga da vrši stanovite procedure pomoću programa za zaključivanje (*inference engine*). Kako bi se to postiglo, bazu znanja treba izgraditi na principu formalnih strojno obradivih struktura podataka. Sam pojam baze znanja zapravo je neformalan naziv za skup informacija u kojem je ontologija jedan od sastavnih dijelova. Međutim, taj skup također može sadržavati i informacije izražene u obliku deklarativnog jezika kao što su logika (*logic*) ili pravila ekspertnog sustava (*expert system rules*), kao i informacije koje nisu formalizirane, već su izražene prirodnim jezikom ili proceduralnim kodom. Ontologija se izgrađuje slijedom temeljne logičke procedure, a rezultat toga je određena klasifikacijska struktura sa jasno određenim kategorijama i pojmovnim odnosima koje se može prikazati putem “pojmovnih grafova” (*conceptual graphs*) i formatizirati na strojno obradiv način (Sowa, 2000).

Prikaz znanja (*knowledge representation*), kako se razumijeva unutar područja umjetne inteligencije, bavi se širokim spektrom računalno izvedivog znanja tj. znanja izraženog putem strogih pravila logike. Primjer takve tvrdnje (*predicate*) “svako A ima 10 s , poneka w su A , poneka w imaju 10 s ” može se formalnije izraziti, odnosno, računalno izvesti na sljedeći način: “za svaki x od A postoji niz od s , pri čemu vrijednost od s iznosi 10, a za svako w u s , w je A , a x sadrži w ”.

Izražajni potencijal logike obuhvaća svaku vrstu informacije koja se može pohraniti ili programirati na računalu. Međutim logika sama po sebi ne može opisati postojeće elemente stvarnosti. Sama logika je jednostavan jezik sa malim brojem osnovnih znakova, pa prema tome temeljne tvrdnje koje bi predstavljale znanje o stvarnom svijetu (u navedenom primjeru su prikazane znakovima x , A , s , w) moraju proizaći iz ontologije. Ontologija se izgrađuje slijedenjem temeljne logičke procedure, a rezultat toga je određena klasifikacijska struktura sa jasno određenim kategorijama i pojmovnim odnosima koje se može prikazati, na primjer, kroz formalizirane strukture koje se nazivaju pojmovni

grafovi (*conceptual graphs*) i formatizirati na strojno obradiv način (vidi Sliku 1)(Sowa, 2000)³.



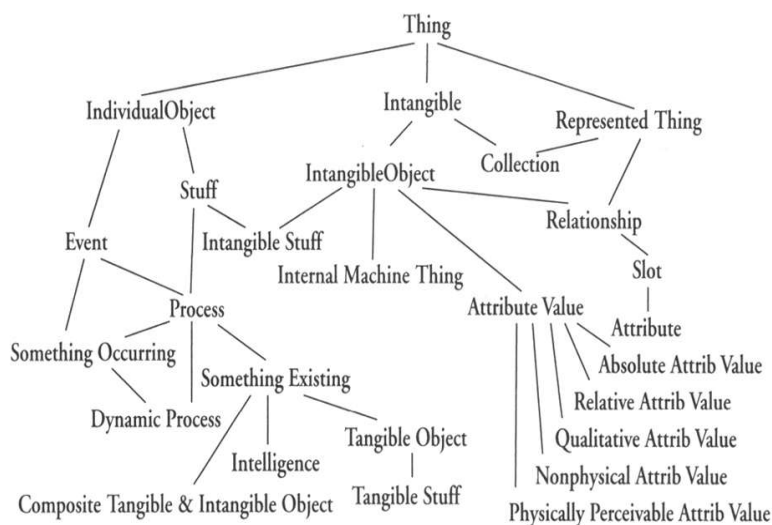
Slika 1: Grafički prikaz (semantička mreža) tipične ontologije “mikrosvijeta” (Sowa, 2000:53)

Ontologijske kategorije grade se promatranjem i promišljanjem na osnovu informacije o apstraktnim i konkretnim entitetima, te vrstama njihovih odnosa unutar određenog područja. Tako ontologiju možemo definirati kao disciplinu koja istražuje kategorije postojećih elemenata stvarnosti ili elemenata stvarnosti koji bi mogli postojati unutar područja. Rezultat takvog istraživanja je katalog koji sadrži vrste elemenata stvarnosti koji postoje ili se za njih pretpostavlja da bi mogli postojati unutar dogovorenog područja, a gledano iz perspektive osobe koja se u opisu područja služi dogovorenim jezikom. To znanje o fizičkom svijetu služi za stvaranje šireg okvira apstrakcija i nudi temeljne postavke i njihove numeričke vrijednosti koji su potrebni da bi ih se moglo izračunavati. Temeljne pretpostavke u ontologiji su ili bazirane na pojedinim područjima ljudskog djelovanja što znači da njihova primjena nije univerzalna ili su neovisne u odnosu na specifična područja i predstavljaju opće primjenjiva svojstva (npr. dio, skup, numerička vrijednost, matematički izraz, itd.). Spajanjem logike i ontologije nastaje jezik kojim je moguće iskazati odnose između entiteta određenog interesnog područja.

³ Neki autori s područja umjetne inteligencije smatraju da ontologija nije naprosto katalog svijeta, taksonomija ili terminologija sama po sebi, već da predstavlja širi okvir u kojem se mogu organizirati taksonomije, terminologije, itd. Guarino i Sowa, međutim, drže da se ontologija nužno mora sastojati od specifičnih pojmova (Guarino, 1997; Sowa, 2000)

Ontologije se međusobno razlikuju po širini pristupa, razini formalnosti, specifičnosti i s obzirom na namjenu. Postoji nekoliko različitih kriterija na temelju kojih se može opisati neka ontologija. Gledano s aspekta područja koje pokrivaju, pojedine ontologije se, poput bibliografskih klasifikacijskih sustava, mogu svrstati u one koje se bave znanjem ograničenim na specifično područje ili one koje obuhvaćaju znanje u njegovom najširem smislu (vidi opširnije Ding, 2001; Sowa, 2000).

Dok filozofi sastavljaju ontologije polazeći od vrha prema dnu, unutar područja računarstva ontologija se obično gradi u obrnutom smjeru. Računalne ontologije često polaze od “mikrosvjetova” koji se lako mogu analizirati, oblikovati i primijeniti. Tako se kod odabira odgovarajućih ontologijskih kategorija može koristiti bilo koji skup kategorija koji se može prikazati u računalnoj primjeni: modeli odnosa između pojedinih entiteta (*entity relationship models – ER*) unutar baze podataka, skup kategorijskih definicija unutar objektnog sustava, geografski pojmovi nužni za određene programske primjene, itd. Ontologije koje služe razmjene znanja među različitim sustavima moraju se bazirati na općenitijim pojmovnim kategorijama. Takva vrsta ontologije ima isti cilj i zahtijeva filozofsku pozadinu sličnu općim ili enciklopedijska bibliografskim klasifikacijama (usporedi Sliku 1 i Sliku 2).



Slika 2: Grafički prikaz pojmovnog okvira ontologije Cyc – tip opće ontologije (Sowa, 2000:54)

Unutar konteksta semantičkog Weba pojam ontologije može imati vrlo široko značenje koje se obično izvodi iz klasifikacijske strukture i sustava upravljanja pojmovnim rječnikom koji su sastavni dijelovi svake ontologije. McGuin-

ness (2002) nudi kategorizaciju u smislu praktične primjene ontologija koja ilustrira sličnost s, e.g. bibliografskom klasifikacijom, a koja dijeli ontologije na **jednostavne i strukturirane**.

Govoreći o **jednostavnim ontologijama** McGuinness daje primjere taksonomija ili jednostavnih hijerarhijskih rječnika⁴ i definira njihovu svrhu na način koje naglašava njihovu sličnost s knjižničnim klasifikacijama :

- osiguravaju kontrolu rječnika
- služe organizaciji mrežnih čvorova i navigaciji
- služe kao predmetno krovište u proširenju sadržaja
- omogućuju predmetno pregledavanje
- omogućuju predmetno pretraživanje
- razrješavaju dvosmislenost u terminologiji

Strukturirane ontologije⁵, s druge strane, osim strojno čitljivih, kodiranih hijerarhijskih odnosa, sadrže informacije o pojedinim svojstvima i njihovim vrijednosnim ograničenjima čime se određeni pojam izravno povezuje sa konkretnim primjerom na kojeg se odnosi. Na primjer, kategorija “proizvod” može imati svojstvo “cijena” čija se vrijednost može ograničiti brojem ili skupom brojeva. Budući se pojmovi unutar neke ontologije iskazuju kroz kategorije, svojstva i funkcije koje su kodirane na stroju čitljiv način, bilo koji dio kodirane strukture nekog pojma može se nadalje pobliže definirati kroz vrijednosti koje se s tim dijelom mogu povezati. U tom smislu strukturirane ontologije, se mogu koristiti kao dio računalne primjene gdje služe za:

- provjeravanja dosljednosti
- izvršavanje (određeno svojstvo automatski uključuje/isključuje druga svojstva)
- podršku komunikaciji s ostalim sustavima (do informacija koje nedostaju može se doći putem veze sa ostalim svojstvima)
- kodiranje programskih setova za testiranje
- podršku konfiguraciji (informacije o sustavima na koji se ontologija primjenjuje)

⁴ Primjer: DMOZ <http://www.dmoz.com> i UMLS – Unified Medical Language <http://www.nlm.nih.gov/research/umls>

⁵ Primjer: WorldNet (strukturirani vokabular engleskog jezika); Cyc (opća univerzalna ontologija); SENSUS (ontologija iz lingvističke domene) i STEP (ontologija sastavljena sa svrhom omogućavanja razmjene podataka o proizvodima između različitih računalnih sustava).

- strukturirano i prilagodljivo pretraživanje
- korištenje općih i specijalnih informacija

2.2 Jezici za formatiranje ontologije

Da bi bila strojno čitljiva ontologija mora biti prikazana jezikom koji pruža sustav kodiranja pogodan za strojnu obradu (*representation language*). Jezik za kodiranje ontologije koja služi ekspertnom sustavu baziranom na složenom ontološkom okviru, bazi znanja i logičkim pravilima, bit će, zasigurno, mnogo kompleksniji u odnosu na jezik kojim se, na primjer, izražavaju jednostavni taksonomijski odnosi među pojmovima. Novija istraživanja usredotočuju se na uspostavljanje potrebnih standarda na tom području pri čemu se izražajni potencijal jezika nastoji spojiti sa sposobnošću logičkog razmišljanja kako bi se dobio jedan sveobuhvatan jezik za reprezentaciju znanja s visokim stupanjem djelatne unutarnje logike (McGuinness, 2002). U području softverskog inženjeringa postoje i neki alternativni pristupi oblikovanju ontologija utemeljenih na konstruktima modeliranja, analizi i dizajnu objektno-orijentiranih programskih primjena (Cranefield, 2001).

Jezici koji se koriste u prikazu ontologija pripadaju trima kategorijama: jezici koji se temelje na logici prvog reda (*first-order logic*)⁶, jezici koji se temelje na logici sustava okvira (*frame logic*)⁷ i mrežni jezici (RDF, XML, HTML). Dok su prve dvije kategorije u potpunosti zadane konkretnim primjenama umjetne inteligencije, mrežni jezici za prikaz ontologija mogu se koristiti u prikazu rječnika za organizaciju znanja izvan domene umjetne inteligencije.

Jedan od najznačajnijih pomaka na području ontologijskih mrežnih jezika je DAML (*DAML – DARPA Agent Markup Language*) koji je posebno osmišljen kao jezik i sredstvo za ostvarenje ideje semantičkog Weba. DAML je karakterističan primjer standarda koji se služi web-kompatibilnim jezikom zajedno sa logičkom paradigmom iz područja umjetne inteligencije. Najnovija verzija tog jezika (DAML + OIL) sastoji se od bogatog niza konstruktivnih elemenata pogodnih za sastavljanje ontologija i detaljno označavanje informacija tako da one budu čitljive i razumljive stroju. U njemu su spojeni mrežni jezik, deskriptivna logika i sustav rasuđivanja utemeljenog na principu okvira sa slojevitim pristupom izgradnji jezika kako ju propisuje OIL (*The Ontology Inference Layer*).

⁶ Tipičan primjer je KIF (Knowledge Interchange Format) služi uspostavljanju veze između ontologije i programskog jezika i nudi KIF specifičnu sintaksu za izražavanje predikatne logike. Temeljna svrha KIFa, kojeg karakterizira visok stupanj računalne složenosti, je da služi kao jezik za razmjenu između raznorodnih baza znanja i baza podataka.

⁷ Primjeri ove vrste ontologijskih jezika su: KL-ONE (Knowledge Language One) koji podržava okvirnu strukturu podataka sa mrežnim sustavom notacije i jasno određenom semantikom, FRL (Frame Representational Language) koji podržava okvirnu strukturu podataka koja se koristi u prikazu stereotipnih situacija.

DAML je uskladio sa RDF shemom (RDFS) i sadržava preciznu semantiku za opisivanje značenja pojmova (jednako tako pogodnu i za opisivanje informacija koje se podrazumijevaju) (Fensel, 2000).

OWL (*Web Ontology Language*) je jezik za semantičko kodiranje (*semantic mark-up*) razvijen od strane W3C. OWL je izveden iz DAML standarda u svrhu kreiranja i razmjene ontologija na Webu. (<http://www.w3.org/TR/owl-ref/>). Standard je namijenjen programskim aplikacijama, koje umjesto jednostavne prezentacije informacija za ljudsko korištenje, trebaju obrađivati sadržaj informacija. OWL omogućava strojnu razmjenu Web sadržaja na razini koja je naprednija od one podržane jednostavnim XML, RDF ili RDFS shemama. OWL ima tree 'podjezika' različitog stupnja ekspresivnosti: *OWL Lite*, *OWL DL* i *OWL Full*.

Još jedan primjer ontologijskog mrežnih jezika je SHOE (*Simple HTML Ontology extension*) Radi se o jeziku za prikaz znanja baziranom na HTML-u koji omogućava kodiranje kategorija, relacija, atributa, pravila za zaključivanje unutar ontologije. SHOE pruža relativno visoku razinu iskazivanja semantike te omogućava web dizajnerima opremanje dokumenata informacijama o općem "sadržaju" ali i bilo kojoj specifičnoj informaciji unutar samog dokumenta. Nadalje, SHOE olakšava automatskim programima za pretraživanje Web-a (*intelligent agents*) automatsko izvođenje zaključaka o podacima koje prikupljaju, hijerarhijsku organizaciju informacijskih te visoko razvijeni ontologijski mehanizam posebno prilagođen potrebama mreže (<http://www.cs.umd.edu/projects/plus/SHOE/>).

3. Sustav za organizaciju znanja i mrežno okruženje

Razvoj u području ontologijskih mrežnih jezika ima značajne posljedice po buduće korištenje tradicionalnih sustava za organizaciju znanja, kao što su klasifikacijski sustavi ili tezaursi, u mrežnom okruženju. Najvažnija karakteristika neke ontologije je njena temeljna klasifikacijska struktura ili taksonomija. Oblikovanje te strukture unutar ontologije mora slijediti stroga pravila logike u klasifikaciji pojmova koja striktno razgraničavaju čiste taksnomijske odnose (tj. podgrupiranje rod/vrsta) od formiranja razreda pojmova na osnovu partitivnih odnosa, instancija i jezičnih višeznačnica. Guarino i Welty (2002) ukazuju, primjerice, na tipične greške u klasifikaciji pojmova koje ontologija ne može tolerirati. Autori navode primjere pogrešne diobe u kojima sve instancije razreda nisu nužno instancije nadrazreda, ili bitna svojstva (*essential properties*) nisu zajednička razredu i svim podrazredima, ili se podjela prema svojstvu koja ne određuju suštinu dalje razrađuje prema suštinskim svojstvima itd.

Ako pogledamo postojeće klasifikacijske sustave ustanoviti ćemo da među njima postoje vrlo velike razlike s obzirom na razinu formalnosti u strukturira-

nju i prikazu podataka i očigledno je potrebno uložiti dodatni napor unutar svakog od tih sustava kako bi njihova semantika i sintaksa bile jasnije definirane, tj. transparentnije (Cordeiro & Slavic, 2002; Slavic & Cordeiro, 2004). Uobičajeno je da se proces konverzije indeksnih jezika u strojno čitljivi oblik temelji na modelu koji se sastoji od vokabulara, sintakse, semantike i izvedbenih pravila tog specifičnog sustava. Taj se proces pokušava dodatno standardizirati kako bi se postojeće sustave za organizaciju znanja što bolje iskoristilo u razvoju semantičkog Weba (Soergel, 1999; Soergel, 2001; Gilchrist, 2002). Cilj koji bi se dobro nadopunjavao s idejom semantičkog Weba je da se osigura široka dostupnost svih sustava za organizaciju znanja putem Weba, u strojno-razumljivom obliku i to na način da svaki indeksni pojam bude jedinstveno prepoznatljiv, ali istovremeno objašnjen kroz strukturu sustava kojem pripada.

Već su učinjeni važni pomaci u pravcu izgradnje tzv. *terminoloških službi* (terminological services) kako bi se uspostavili opći standardi za prikaz i razmjenu sustava za organizaciju znanja dio kojih su tradicionalni indeksni jezici. Preduvjet tog procesa su otvoreni i univerzalni standardi kodiranja kontroliranih rječnika. Soergel (2001:1) ovako objašnjava svrhu tih standarda:

- i) mogućnost unošenja podataka iz sustava za organizaciju znanja u program i njihovog prenošenja iz jednog programa u drugi
- ii) mogućnost pristupa sustavima za organizaciju znanja da bi se pronašlo primjere primjene i mogućnost postavljanja pitanja i dobivanja odgovora unutar sustava
- iii) mogućnost prepoznavanja specifičnih pojmova unutar pojedinog sustava za organizaciju znanja kojemu ti termini i pojmovi pripadaju
- iv) propisivanje i upućivanje na primjere uspješne primjene

Napredak unutar šire Internetom povezane zajednice uključuje napore da se postojeći podaci iskoriste što bolje i to iznalaženjem zajedničke osnove, uopćenih sustava i oblika prikaza i protokola za povezivanje i razmjenu različitih vrsta indeksnih jezika (Qin & Paling, 2001; Vizine-Goetz, 2003; Vizine-Goetz et al., 2004; Zeng & Lois, 2004). Jedna od aktivnih inicijativa na ovom području je forum NKOS – *Networked Knowledge Organization Systems/Services* koji djeluje putem radionica, publikacija i elektroničke diskusijske liste i na taj način povezuje istraživače koji rade na praktičnoj primjeni i one koji se bave razvijanjem standarda unutar pojedinih profesionalnih područja (<http://nkos.slis.kent.edu/>).

3.1 Standardi za korištenje i razmjenu sustava za organizaciju znanja

Standardi zasebnih informacijskih sektora i domena (knjižničarstvo, digitalne knjižnice, geografske i geospacijalne informacije, informacijski sustavi

državne uprave i vlada, arhivi, sustavi za elektroničko obrazovanje) sada se testiraju primjenom transparentnijih i fleksibilnijih standarda za prijenos podataka kao što je, na primjer, XML ili XML/RDF. Tipičan su primjer standardi za strojnu prevođenje i strojnu razmjenu rječnika, glosara, pojmovnih mapa, kao i o već razvijeni standardi za razmjenu tezaurusa i klasifikacijskih sustava. Neki od ovih standarda stvaraju se imajući u vidu potrebe semantičkog Web-a.

Standardi za razmjenu terminologije (rječnika, glosara)

OLIF – Open Lexicon Interchange Format je format uskladiv sa XML-om i jednostavan je za korištenje. Služi za razmjenu terminoloških i leksičkih podataka. Standardi leksičke i terminološke razmjene trenutno se razvijaju unutar OLIF Konzorcija, kooperativne skupine industrijskih tvrtki koje se angažiraju na području jezične tehnologije (<http://www.olif.net/>)

MARTIF – ISO 12200:1999 – Machine-Readable Terminology Interchange Format – MARTIF je na SGML-u utemeljen format za razmjenu podataka između konceptualno orijentiranih terminoloških baza podataka. MARTIF je namijenjen razmjeni među partnerima (npr., dva prevodilačka poduzeća) koji se “poznaju” i mogu “pregovarati” oko pojedinosti formata kako bi gubitak informacija sveli na najmanju moguću mjeru.

TMF – ISO 16642:2003 – Terminological Markup Framework, je utemeljen na OLIF-u i MARTIF-u koji se usredotočuju uglavnom na leksičke podatke. TMF je, međutim, više mrežno orijentiran i uključuje svojstva pogodna da obuhvati pojmovnih i ontoloških aspekata terminoloških podataka (Romary, 2001). TMF pobliže određuje temeljne strukture i mehanizme nužne za računalni prikaz terminologija, bez obzira na to o kojem se specifičnom formatu radi. Smisao ovog standarda je da ukaže na ograničenja prikaza računalno obrađene terminologije i da održi sukladnost između prikaza. Konkretni primjer reprezentacijskog formata koji je izveden iz TMF-a je *Terminological Mark-up Language* (<http://www.loria.fr/projects/TMF/>).

Standardi za knjižnične sustave za organizaciju znanja

MARC Authority Formats – Formativni normativni datoteka iz MARC skupine formata podržavaju normativnu kontrolu klasifikacijskih sustava. Njihova primjena je u velikoj mjeri ograničena postojećim standardom za kodiranje bibliografskih formata i za sada ne postoji nikakav alternativni tip mrežno orijentiranog kodiranja. MARC formati posebno su namijenjeni za primjene knjižničnih sustava u kontekstu specifičnih klasifikacijskih sustava, a sa stanovišta oblikovanja podataka pogodniji su za primjene u knjižničnim integriranim sustavima nego za mrežno okruženje. Među njima valja istaknuti *MARC 21 Concise*

Format for Classification Data (http://www.loc.gov/marc/classification/eccd_home.html) i *Concise UNIMARC Format for Classification Data* (<http://www.ifla.org/VI/3/p1996-1/concise.htm>).

Zthes profile – je protokol tipa Z39.50 za pregledavanje tezaurusa. Opisuje apstraktni model za prikaz i pretraživanje tezaurusa (npr. hijerarhije termina kako su opisane u ISO 2788: 1986) i pobliže označava kako se taj model može praktično primijeniti korištenjem protokola Z39.50. Također ukazuje na moguće primjene ovog modela kroz korištenje nekih drugih protokola i formata. Kao dodatak profilu izrađen je ZThes DTD za primjenu rječnika formatiranih u XML.

Standardi za rječnike u domeni elektroničkog obrazovanja

VDEX – Vocabulary definition exchange je specifikacija kreirana u okviru IMS Global Learning Consortium kojom je definirana gramatika za razmjenu terminologije različitog tipa koje se koriste u metapodacima obrazovnog materijala. VDEX se može koristiti u razmjeni jednostavnih strojno čitljivih popisa pojmova ili pojmova zajedno s nekim dodatnim informacijama koje ljudskim korisnicima mogu znatno olakšati razumijevanje značenja ili konkretne primjene raznih pojmova. VDEX XML sintaksa primjenjuje se na sheme sa striktno definiranim hijerarhijama. VDEX Version 1 Final Specification je inačica standarda službeno odobrena od strane IMS-a u veljači 2004.

Opći standardi za razmjenu rječnika

XTM Topic Maps – ISO/IEC 13250:2000 nudi specifikaciju i gramatiku primjenjivu u prikazu izvora informacija koji se koristi pri definiranju tema (topics) i relacija među temama. Mrežno-orijentirana verzija standarda nazvana *XML Topic maps (XTM)* je skup standarda razvijen s ciljem pojednostavljivanja primjene paradigme tematskih struktura na webu i iskorištavanju potencijala za pronalaženje i upravljanje informacijama (<http://www.topicmaps.org/xtm/1.0/>). Tematska svojstva zadana su nekim okvirima, tj. ograničenim kontekstima unutar kojih se nazive i informacijske izvore smatra nazivima, izvorima i tipovima odnosa koji pripadaju toj konkretnoj temi. Dokument ili više međusobno povezanih dokumenata u kojima je prisutan takav tip gramatike naziva se “tematska karta” (Sigel, 2004).

Voc-ML – The Vocabulary Markup Language je standard za izradu opisa pojedinih sustava za organizaciju znanja koji je nastao je kao rezultat zaključaka NISO radionice održane 1999. na temu “Elektronički tezaurusi: osmišljavanje i planiranje standarda” pri čemu je ustanovljeno da se svaki pojedinačni kontrolirani rječnik mora moći opisati i identificirati kao cjelina (Milstead, 1999).

NKOS je preuzeo inicijativu i definirao skup svojstava nužnih za opis kontroliranih rječnika te predložio prvu verziju XML DTD nazvanu Vocabulary Markup Language (Vocabulary ML, 2000). Voc-ML uključuje Dublin Core metapodatke koje sadrže opis sustava koji se enkodira. Standard također definira oznake i sintaksu za jedinstveno označavanje svakog pojma, njegovog odnosa naspram drugih pojmova, kao i opisnih informacija kao što su napomene o kontekstu i definicije. Očekuje se da će ova shema u svom konačnom obliku biti primjenjiva na niz sustava različitih vrsta, uključujući normativne datoteke, hijerarhijske tezaure, klasifikacijske sheme, digitalne geografske leksikone i predmetne odrednice (Hodge, 2000).

XFML – Exchangable Faceted Metadata Language služi upotrebi i razmjeni facetnih kontroliranih rječnika (<http://www.xfml.org/spec/1.0.html>). U suštini, XFML je, poput XTM standarda, paradigma za izražavanje tema u formi hijerarhija korištenjem “spremnika” za organizaciju međusobno isključivih kategorija pojmova (faceta) koji omogućuje da se cjelokupni sadržaj objavi u XML formatu. Postoji mogućnost povezivanja različitih XFML tematskih karti kroz utvrđivanje istovjetnosti pojmova koje pripadaju različitim strukturama (Van Dijk, 2003).

SKOS – Knowledge Organization Systems and Semantic Web se odnosi na specifikaciju i standarde koji bi olakšali upotrebu sustava za organizaciju znanja (KOS) na semantičkom Webu. Na razvoju SKOS-a rade članovi radne grupe “W3C SWBP-WG Tezaurus” u okviru europskog projekta SWAD. SKOS se koristi fleksibilnom XML/RDF sintaksom i namijenjen je ne samo tezaursima, već i taksonomijama, glosarima, web direktorijima, itd. Trebao bi se koristiti kao ontologija i nadopunjava se sa OWL jezikom (*Ontology Web Language*). Središnja struktura SKOS-a pruža temelj za objavljivanje pojmova sustava za organizaciju znanja i njihovih međusobnih odnosa kako bi se olakšalo pretraživanje i opći pregled, te kako bi se omogućilo dodatno mapiranje i povezivanje pojedinih sustava za organizaciju znanja. 2004. godine SKOS je službeno proglašen “otvorenim” razvojnim procesom (<http://www.w3.org/2004/02/skos/>).

4. Zaključno

Pretraživanje informacija na Internetu naglasilo je potrebu za predmetnom organizacijom sadržaja i njihovog pronalaženja što je oživjelo interes za tradicionalne sustave za organizaciju znanja (naročito za klasifikacijske sustave i tezaure). Daljnji razvoj ideje semantičkog Weba ovisi o infrastrukturi metapodataka koju se može slikovito prikazati kao prijenosni mehanizam i ‘pogon’ u primjeni indeksnih jezika. Trenutno se radi na standardiziranju formata za razmjenu rječnika koji će se koristiti u metapodacima izvora informacija kako bi ih strojevi lakše obrađivali, povezivali i razmjenjivali. Vrijednost tradicionalnih

sustava za organizaciju znanja bit će određena njihovom strojnom čitljivošću i mogućnosti da budu objavljeni i razmjenjivani u otvorenom mrežnom okruženju.

Mrežno okruženje daje prednost globalnim rješenjima koja su neovisna o programskoj platformi i lokalnim programskim rješenjima te na taj način nameće filozofiju otvorenog informacijskog prostora u kojem se jedinstveno tehnološko rješenje koristi za prijenos različitih vrsta sadržaja. Jedan od temeljnih problema unutar arhitekture metapodataka i temeljne infrastrukture semantičkog Weba odnosi se na postavljanje jedinstvenih trajnih identifikatora za sheme različitih rječnika kako bi se pojedini pojmovi mogli nepogrešivo navoditi u neograničenom broju zapisa metapodataka. Pomaci u razvijanju standarda za detaljno označavanje rječnika i mrežnih ontologijskih jezika su usklađivi sa arhitekturom Weba unutar koje bi stroju razumljivi tezaursi i klasifikacijske sheme trebali biti dostupni i razmjenjivi preko Weba. Da bi to bilo tako svaki bi pojam koji pripada nekom kontroliranom rječniku u ovom kontekstu morao imati jedinstveni identifikator koji bi omogućio da ga se poveže, semantički protumači, višestruko iskoristi, prenosi i razmjenjuje u različitim shemama metapodataka (Van de Sompel i grupa autora, 2004; Vocabulary ML, 2000; vidi također web stranicu NKOS-a na <http://nkos.slis.kent.edu/>). Trenutno se razmatraju mogući pristupi rječničkim repozitorijima (*vocabulary registries*), tj. sustavu za registraciju rječnika i terminološkim službama (*terminological services*) (Pepper i Garshol, 2002; Vizine-Goetz, 2003; Vizine-Goetz et al., 2004).

Bitan pomak na području identifikacije pojedinih pojmova unutar vokabulara sastoji se, u još nedovoljno, razrađenoj ideji identifikacije nazvanoj PSI – *Published Subjects Identifier*: PSI je otvoreni, distribuirani mehanizam za definiranje jedinstvenih globalnih identifikatora. Utemeljen na URI, ovaj mehanizam ima dvije jedinstvene karakteristike: način funkcioniranja od dna prema vrhu, te činjenicu da ga mogu koristiti i ljudi i strojevi” (*Published subjects: introduction and basic requirements*, 2003).

Upotreba klasifikacijskih i drugih tradicionalnih sustava u metapodacima i općenito u organizaciji pronalaznja izvora informacija na Internetu uvelike će ovisiti o tome u kojem će obliku kontrolirani rječnici biti dostupni. Kao što ističu Vizine-Goetz et al. (2004), za uspostavljanje terminoloških službi nužno je unaprijediti i znatno poboljšati postojeće formate kontroliranih rječnika te ih opremiti trajnim identifikatorima i podacima o porijeklu, relacijama i razvoju pojmova. U ovom trenutku sustavi za organizaciju znanja i praktično znanje o njima nisu dostupni u obliku pogodnom za široku upotrebu na Internetu, a i infrastruktura nužna za razmjenu pomagala tradicionalne organizacije znanja nije još posve spremna za upotrebu izvan bibliografske domene (Cordeiro & Slavić, 2002; Slavić i Cordeiro, 2004; Slavić & Cordeiro, 2004a).

Literatura

- BERNERS-LEE, T. (1996) "The World Wide Web – past, present and future. Presentation at the British Computer Society on July 17, 1996. <http://jodi.ecs.soton.ac.uk/Articles/v01/i01/BernersLee/>.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. (2001) "The semantic web", *Scientific American*, May 2001. [Also available at <http://www.sciam.com/2001/0501issue/0501berners-lee.html>].
- BROOKS, T. A. (2001) "Where is meaning when form is gone?: knowledge representation", *Information Research*, 6 (2) 2001. <http://informationr.net/ir/6-2/paper92a.html>.
- CORDEIRO, M. I.; SLAVIC, A. (2002) "Data models for knowledge organization tools: evolution and perspectives", *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries: proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain*. Eds. María J. López-Huertas with the assistance of Francisco J. Muñoz-Fernández. Würzburg: Ergon Verlag, 2002. (Advances in Knowledge Organization 8), 127-134.
- CRANEFIELD, S. (2001) "Networked knowledge representation and exchange using UML and RDF", *Journal of Digital Information*, 1 (8) 2001. <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Cranefield/>.
- CROFT, B. W. (1997) "What do people want from information retrieval", *From classification to "knowledge organisation": Dorking revisited or "past is prelude": a collection of reprints to commemorate the forty year span between the Dorking Conference (First International Study Conference on Classification Research 1957) and the Sixth International Study Conference on Classification Research (London, UK) 1997*. Edited by Alan Gilchrist. The Hague: International Federation for Information and Documentation (FID), 1997. (FID 714), 181-185.
- DING, Y. (2001) "A review of ontologies with the semantic web in view", *Journal of Information Science*, 27 (6) 2001, 377-384.
- FENSEL, D. et al.: (2000) "OIL in a nutshell", *Knowledge Acquisition, Modeling, and Management: proceedings of the European Knowledge Acquisition Conference (EKAW-2000)*. R. Dieng et al. (eds.). Würzburg: Springer-Verlag, 2000.
- GILCHRIST, A. (2002) "From Aristotle to the 'semantic web'", *LA Record*, 104 (1) 2002. <http://www.la-hq.org.uk/directory/record/r200201/article2.html>
- GILCHRIST, A. (2003) "Thesauri, taxonomies and ontologies – an etymological note", *Journal of Documentation*, 59 (1) 2003, 7-18.

- GUARINO, N. (1997) "Understanding, building and using ontologies", *International Journal of Human-Computer Studies*, 46 1997, 293-310.
[Also available at <http://www.ladseb.pd.cnr.it/infor/Ontology/>]
- GUARINO, N.; WELTY, C. (2002) "Evaluating ontological decisions with OntoClean", *Communications of the ACM*, 45 (2) 2002, 61-65.
- HODGE, G. (2000) "Systems of knowledge organization for digital libraries: beyond traditional authority files". Council of Library and Information Resources, April 2000, <http://www.clir.org/pubs/reports/pub91/contents.html>.
- LASSILA, O.; MCGUINNESS, D. (2001) "The role of frame-based representation on the Semantic Web", *Linköping Electronic Articles in Computer and Information Science*, 6 (5) 2001, <http://www.ida.liu.se/ext/epa/cis/2001/005/paper.pdf>.
- MCGUINNESS, D. L. (2002) "Ontologies come of age", *Spinning the semantic web: bringing the World Wide Web to its full potential*. Eds. D. Fensel et. al. Cambridge, MA: MIT Press, 2002, 190. [Also available at [http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-\(with-citation\).htm](http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm)].
- NOY, N. F.; MCGUINNESS, D. L. (2001) "Ontology development 101: a guide to creating your first ontology", Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001. http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html.
- PEPPER, S. ; GARSHOL, L. M. (2002) "Lessons on Applying Topic Maps", [paper presented at] *XML Conference and Exposition, Baltimore MD, December 8-13 2002*. <http://www.ontopia.net/topicmaps/materials/xmlconf.html#XML-Europe-2002>.
- Published Subjects: introduction and basic requirements: [status: committee specification]. (2004) OASIS [The Organization for the Advancement of Structured Information Standards] Published Subjects Technical Committee Recommendation, 2003-06-24. Editor Steve Pepper. <http://www.oasis-open.org/committees/download.php/3050/pubsub-pt1-1.02-cs.pdf>.
- QIN, J.; PALING, S. (2001) "Converting a controlled vocabulary into an ontology: the case of GEM", *Information Research*, 6 (2) 2001. <http://informationr.net/ir/6-2/paper94.html>
- ROMARY, L ; CAMPENHOUDT, M. Van. (2001) "Normalisation des échanges de données en terminologie: le cas des relations dites 'conceptuelles'", [paper presented at] *Conférence TIA-2001, Nancy, 3 et 4 mai 2001*. <http://www.termisti.refer.org/tia4.pdf>.

- SCHMITZ-ESSER, W. "How to cope with dynamism in ontologies", Dynamism and stability of knowledge organization: proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada, edited by Clare Beghtol, Lynne C. Howarth, Nancy J. Williamson. Würzburg: ERGON Verlag, 2000, (Advances in knowledge organization 7), 83-89.
- SIGEL, A. (2004) "Wissensorganisation, Topic Maps und Ontology Engineering: Die Verbindung bewährter Begriffsstrukturen mit aktueller XML-Technologie 1", [presented at 7. Deutschen ISKOTagung in Berlin, March 2004 and will be published in Tagungsband der 7. Deutschen ISKO-Tagung, Würzburg: Ergon Verlag]. http://kpeer.wim.uni-koeln.de/~sigel/veroeff/ISKO-2001/ISKO2001_Berlin_SIGEL_v24.pdf.
- SLAVIC, A.; CORDEIRO, M. I. (2004) "Core requirements for automation of analytico-synthetic classifications", *Knowledge Organization and the Global Information Society: proceeding of the 8th ISKO Conference, London 13-16 July 2004*. Edited by I. C. McIlwaine. Würzburg: Ergon Verlag, 2004. (Advances in Knowledge Organization 9), 187-192.
- SLAVIC, A.; CORDEIRO, M. I. (2004a) "Sharing and re-use of classification systems: the need for a common data model", *Signum*, 2004 [u tisku].
- SOERGEL, D. (1999) "The rise of ontologies or the reinvention of classification", *Journal of the American Society for Information Science*, 50 (12) 1999, 1119-1120.
- SOERGEL, D. (2001) "The representation of Knowledge Organization Structure (KOS) data: a multiplicity of standards", [paper presented at] *JCDL NKOS Workshop, VA, 2001*. <http://www.dsoergel.com/cv/B75.pdf>.
- SOWA, J. F. Knowledge representation: logical, philosophical and computational foundations. Pacific Grove, CA: Brooks/Cole, 2000.
- VAN DIJK, P. (2003) "eXchangeable Faceted Metadata Language – XFML Core. [Last changed in 2003]. <http://www.xfml.org/spec/1.0.html>.
- VATANT, B. (2004) Published subject nocturne at XML 2004. 2004-11-09. [message to the mailing list: public-esw-thes@w3.org]. <http://esw.w3.org/mt/esw/archives/000094.html>.
- VELTMAN, K. H. (2001) "Syntactic and semantic interoperability: new approaches to knowledge and the semantic web", *The New Review of Information Networking*, 7 2001, 159-183.
- VELTMAN, K. H. (2002) "Challenges for a Semantic Web", *Proceedings of the International Workshop on the Semantic Web 2002 (at the Eleventh International World Wide Web Conference), Honolulu, Hawaii, May 7, 2002*.

<http://semanticweb2002.aifb.uni-karlsruhe.de/proceedings/Position/veltmann.pdf>.

VELTMAN, K. H. (2004) "Towards a Semantic Web for Culture", *Journal of Digital Information*, 4 (4) 2004, article no. 255. <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Veltman/>.

VICKERY, B. C (1997) "Ontologies", *Journal of Information Science*, 23 (4) 1997, 277-286.

VIZINE-GOETZ, D. (2003) "Terminology services", *Presentation at ECDL conference, 21 August, 2003*. http://staff.oclc.org/~vizine/ECDL2003NKOS/ECDL03NKOS_vizine-goetz.ppt.

VIZINE-GOETZ, D. et. al. (2004) "Vocabulary Mapping for Terminology Services", *Journal of Digital Information*, 4 (4) 2004, article no. 272. <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Vizine-Goetz/>.

Vocabulary ML (2000) : Metacode strawman DTD. NKOS, The Vocabulary Mark-up Language. NKOS, 2000. <http://nkos.slis.kent.edu/>

ZENG, M. L.; LOIS, M. C. (2004) "Trends and issues in establishing interoperability among knowledge organization systems", *Journal of the American Society for Information Science and Technology*, 55 (5) 2004, 377-395.

Semantic Web, knowledge organization systems and network standards

AIDA SLAVIĆ

Summary

The Internet is now fast moving into a new era (Internet 2) with a new World Wide Web architecture known as the *Semantic Web*. In this environment the Web will be gradually populated with content with formalized semantics that will enable the automation of content organization and its retrieval (Berners-Lee & Hendler & Lassila, 2001; Lassila & McGuinness, 2001). As implied by its name, this kind of network will assume a higher level of connectivity which is going to be based on resource content and meaning while the information organization will predominantly be automatic i.e. based on *machine to machine* (m2m) information services. This is the reason for Semantic Web to be closely related with the development of ontologies. Traditional knowledge organization systems (KOS) such as classifications and thesauri have been deployed for resource organization and discovery on the Internet and became widespread in subject gateways and resource discovery metadata. KOS tools are likely to become even more important with the Semantic Web providing they can be exposed and shared using ontologically based standards.

Key words: Semantic Web, ontology, knowledge organization systems, controlled vocabularies, vocabulary exchange, network standards, formats.