

Kooperativan rad na dogradnji Srpskog wordneta

Cvetana Krstev, Filološki fakultet, cvetana@matf.bg.ac.yu

Bojana Đorđević, Filološki fakultet, bojana@lingvistika.org

Sanja Antičić, Univerzitetska biblioteka "Svetozar Marković", antonice@unilib.bg.ac.yu

Nevena Ivković-Berček, Biblioteka Teološkog fakulteta u Beogradu,

ivkoput@yahoo.com

Zorica Zorica, Galerija likovne umetnosti poklon zbirka Rajka Mamuzica,

zzorica@eunet.yu

Vesna Crnogorac, Bibliotekarsko društvo Srbije, crvesna@sezampro.yu

Ljiljana Macura, Narodna biblioteka Srbije, ljiljana.macura@nbs.bg.ac.yu

Apstrakt: U ovom radu predstavljamo informatičku leksičku bazu podataka Wordnet koja je postala *de facto* standard za semantičke mreže. Prva ovakva mreža izgrađena za engleski jezik na Prinstonskom univerzitetu u laboratoriji za kognitivne nauke ostala je i najrazvijenija baza podataka ove vrste, ujedno posluživši kao osnov za izgradnju wordneta za mnoge druge jezike. Wordneti izgrađeni u okviru evropskih projekata EuroWordNet i BalkaNet poravnati su sa Prinstonskim wordnetom što omogućava njihovo korišćenje u mnogim višejezičkim primenama. Tako je i razvoj srpskog wordneta otpočeo u okviru BalkaNet projekta, a potom nastavljen kroz volonterski rad i kooperaciju mnogih saradnika čiji se rad ovde predstavlja.

1. Wordnet

Wordnet je vrlo velika leksička baza podataka koja je organizovana preko čvorova i relacija između tih čvorova (Fellbaum, Christiane, ed. 1998). Ovi čvorovi koji se u wordnetu nazivaju sinsetovi, od engleskog *synset* ili *synonymous set* predstavljaju zapravo skupove reči koje u nekom kontekstu imaju isto značenje. Na primer, u engleskom wordnetu jedan sinset čine reči {teacher:1, instructor:1} sa značenjem «osoba čije je zanimanje da predaje». Brojevi koji prate reči u ovom skupu (ovde je to u oba slučaja broj «1») ukazuju da se ovim rečima može izraziti ovo značenje, ali da se možda pomoću njih može izraziti i neko drugo značenje. Zaista, sinset {teacher:2} odgovara značenju «personifikacija apstrakcije koja podučava», kao u primeru «experience is a demanding teacher» (iskustvo je zahtevan učitelj). Iz ovog primera se vidi da iako su reči sinonimi u prvom kontekstu, one to nisu i u drugom, jer se u poslednjem primeru ne može zameniti «teacher» sa «instructor» (*experience is a demanding instructor).

Ova baza je podeljena u delove prema vrstama reči, i to prema imenicama, glagolima, pridevima i prilozima. Imenički deo baze je organizovan kao hijerarhijska mreža imeničkih čvorova koja se uspostavlja na osnovu postojanja relacije podređenosti i nadređenosti između pojmova koje ti čvorovi predstavljaju. Za jedan pojam kažemo da je podređen drugom pojmu ako poseduje sva svojstva koja poseduje i nadređeni pojam, ali ima i neka specifična svojstva. Ovo se može razjasniti na delu hijerarhije kojoj pripada imenički sinset {teacher:1, instructor:1}. Njegov direktni nadređeni sinset je {educator:1, pedagogue:1} (osoba koja vaspitava mlade ljude), čiji je direktni nadređeni pojam {professional:1, professional person:1} (osoba koja se bavi profesijom koja zahteva visoko obrazovanje), čiji je direktni nadređeni pojam {adult:1, grownup:1} (potpuno zrela odrasla osoba), i tako dalje. Iz ovog primera se vidi, da nadređeni pojmovi sve više gube na svojstvima i u tom smislu predstavljaju sve apstraktnije pojmove. Dalje, takoreći svaki od tih nadređenih pojmova ima više podređenih pojmova. Na primer, odrasla osoba se može baviti nekom učenom profesijom

ali može da bude i Katica za sve ({Jack of all trades:1}), dok učena osoba može da vaspitava mlade ljude ali može i da se bavi pravom ili daje pravne savete ({lawyer:1, attorney:1}) i tako dalje.

Relacija podređenosti i nadređenosti svakako nije jedina koja se uspostavlja između pojmova. Iscrpnija analiza pojmova koji se leksikalizuju pomoću imenica data je, na primer, u (Miller, George A., 1990). Neke od tih relacija uspostavljene su i u Prinstonskom wordnetu čime je uspostavljena jedna složena mreža leksikalizovanih pojmova. Relacije koje posle relacije podređen-nadređen prevladaju u wordnetu jesu relacije deo-celina i član-celina. Kao primer uzmimo pojam, predstavljen skupom {warship:1, war vessel:1, combat ship:1} (brod koji je na raspolaganju državi za vođenje rata), koji je «član», tj. u sastavu, *ratne flote*, pojma koji je u Prinstonskom wordnetu predstavljen sinsetom {fleet:4} (grupa ratnih brodova koja se organizuje kao taktička jedinica). *Ratni brod*, s druge strane, kao svoj deo sadrži *brodski top*, tj. {naval gun:1} koji predstavlja vrstu mornaričkog oružja koje je instalirano na ratnom brodu. S njima u vezi je i relacija «sastojak-celina» kojom su, na primer povezani pojmovi {protein:1} (neko jedinjenje iz velike grupe nitrogenski organskih jedinjenja koja su suštinski deo živih ćelija) i {egg:1} (ovalno rasplodno telo ženke ptice koje se koristi u ishrani).

Još jedna važna relacija koja se uspostavlja između imeničkih sinsetova jeste relacija antonimije kojom se povezuju (približno) suprotni pojmovi. Očigledne primere predstavljaju {female:2, female person:1} (osoba koja pripada polu koji može da rađa decu) i {male:2, male person:1} (osoba koja pripada polu koji ne može da rađa decu) ili {sorrow:1} (osećanje velike žalosti povezano sa gubitkom) i {joy:1, joyousness:1, joyfulness:1} (osećanje velike sreće).

Druga značajna grupa relacija uspostavljenih između sinsetova u Prinstonskom wordnetu jesu one koje povezuje pojmove koji se leksikalizuju različitim vrstama reči. Važna relacija koja povezuje imeničke i pridevske sinsetove jeste «biti u stanju-stanje nečega», a jedan primer predstavlja sinset {cleanness:1} (stanje nekoga ili nečega što je čisto) koji je povezan sa pridevskim sinsetom {clean:1} (koji je bez prljavštine ili nečistoće ili ima naviku da bude čist). Relacija antonimije je česta među pridevima, pa je sinset {clean:1} povezan relacijom «skoro suprotan» sa sinsetom {dirty:1, soiled:1, unclean:1} (koji na sebi ima prljavštinu ili nečistoću). Ovaj sinset je, pak, u vezi sa imeničkim sinsetom {dirtiness:1, uncleanness:1} (stanje nekoga ili nečega što nije čisto) preko relacije «biti u stanju-stanje nečega», dok je ovaj sinset opet povezan relacijom «skoro suprotan» sa sinsetom {cleanness:1}. Ako ovome pridodamo i relacije koje se uspostavljaju između glagolskih sinsetova, kao što je relacija «uzrokuje-uzrokovan» koja, na primer, povezuje sinsetove {stand:10; stand up:2; place upright:1} (postaviti u uspravan položaj) i {stand:1; stand up:4} (biti u uspravnom položaju) jasno je da je u Prinstonskom wordnetu uspostavljena gusta mreža između čvorova.

Priroda ovih relacija je različita. Neke od njih su simetrične, kao što je relacija «skoro suprotan», jer ako *A* ima skoro suprotno značenje od *B*, onda i *B* ima skoro suprotno značenje od *A*, dok su druge asimetrične, kao što je relacija «uzrokuje», jer ako *A* uzrokuje *B*, onda *B* ne uzrokuje *A*. Za asimetrične relacije međutim uvek postoji parnjak, što je za relaciju «uzrokuje» relacija «uzrokovan», pa ako *A* uzrokuje *B*, onda je *B* uzrokovan od *A*. Neke od relacija su po prirodi 1-1, kao «skoro suprotan», dok su druge po prirodi «mnogo-1», kao «podređen-nadređen». Naime, za *A* obično postoji najviše jedno *B* koje ima (skoro) suprotno značenje i obrnuto, dok ako je *A* najčešće jedinstveni nadređeni pojam pojmu *B* dotle *B* najčešće ima više podređenih pojmova. S druge strane, relacija «deo-celina» je po prirodi «mnogo-mnogo» jer jedno *A* može biti deo raznih *B* dok *B* u svom sastavu, osim *A* obično ima i druge delove. Na primer, {handle:1, grip:2, handgrip:1, hold:8} (dodatak nekom objektu koji je tako napravljen da se ovaj za njega može uhvatiti da bi se koristio ili premeštao) ulazi u sastav mnogih predmeta: četka za kosu, kofer, tiganj, kišobran i mnogih drugih, a jasno je

da svaki od ovih predmeta osim drške sadrži i druge delove. Ovo razmatranje o prirodi relacija između čvorova od značaja je sa stanovišta realizacije same baze podataka. Da bi se izbegla redundantnost, kod asimetričnih relacija uvek se beleži samo jedna relacija, a ne i njen parnjak, i to ona koja u najvećem broju slučajeva ima jedinstvenu vrednost. Na primer, za pojam *A* se beleži šta je njegov (najščešće jedinstveni) nadređeni pojam *B*, dok se za *B* ne beleži šta su sve njegovi potencijalno mnogobrojni podređeni pojmovi jer se oni mogu posredno izvesti.

Značaj Prinstonskog wordneta (wordnet.princeton.edu) nije samo u tome što predstavlja veliku bazu leksičkih podataka već i u tome što je našao primenu u mnogim oblastima, od kojih su neke automatsko razlikovanje značenja, ekspanzija termina pri pronalaženju informacija ili konstrukcija strukturne reprezentacije sadržaja dokumenta. U stvari, wordnet je postao tako popularan da se skoro može smatrati standardom u obradi prirodnih jezika. Mnoge primene wordneta opisane su u (Fellbaum, Christiane, ed. 1998). Prinstonski wordnet se stalno nadograđuje, a poslednja verzija za operativni sistem Windows je 2.1, puštena marta 2005.

2. Proširivanje Prinstonskog wordneta

S obzirom na značaj Prinstonskog wordneta njegova struktura je u više navrata proširivana dodatnim informacijama koje bi ga mogle učiniti primenljivijim u prirodnojezičkim obradama. Ovde ćemo ukratko opisati dva proširenja koja su od značaja za projekat koji je predmet ovog rada.

Prvo proširenje odnosi se na proširivanje Prinstonskog wordneta semantičkim domenima. Semantički domeni predstavljaju prirodan način da se uspostave semantičke relacije između značenja reči koje bi se mogle uspešno koristiti u raznim domenima obrade prirodnih jezika. Semantički domeni su polja ljudskog interesovanja kao što su *sport*, *ekonomija* ili *politika* od kojih svako ima specifičnu terminologiju i leksičku koherentnost. Korišćenje domena je uobičajeno u lingvistici (za označavanje semantičkih polja), kao i u leksikografiji (za označavanje polja upotrebe).

Prinstonski wordnet je proširen obeležjima domena tako što je (skoro) svaki sinset anotiran bar jednim obeležjem domena koje je izabrano iz skupa od oko dvesta hijerarhijski organizovanih domena. Na primer, sinset {mouse:1} (neki od brojnih malih glodara koji obično liče na umanjene pacove pošto imaju šiljate njuške i male uši na izduženim telima sa mršavim, obično glatkim repovima) pripada domenu *zoology* 'zoologija', dok sinset {mouse:2; computer mouse:1} (elektronski uređaj koji kontroliše koordinate kursora na ekranu; pomera se po ravnoj podlozi) pripada domenu *computer science* 'računarstvo'. Ova nova informacija o domenu dopunjava već postojeće informacije u wordnetu. Jedan domen može da uključi sinsetove koji pripadaju različitim vrstama reči kao i različitim hijerarhijama. Korist više od dodavanja ovog obeležja wordnetu je to što ono ponekad grupiše više značenja jedne reči u homogeni klaster čime se umanjuje višeznačnost reči u wordnetu koja je inače veoma velika jer su značenja vrlo fino razdvojena. Uzmimo za primer imenicu *time* 'vreme' koja se kao takva (dakle, ne u sastavu kompozita) javlja u Prinstonskom wordnetu u osam sinsetova od kojih pet pripada domenu *time period* 'vremenski period'.

Za obeležavanje domena korišćeno je 200 obeležja domena iz Djuijeve decimalne klasifikacije koja imaju hijerarhijsku strukturu drveta. Na vrhu ove hijerarhije domena su *doctrines*, *free_time*, *applied_science*, *pure_science*, *social_science* i *factotum*, pri čemu je obeležje *factotum* korišćeno tamo gde se ni jedan drugi domen nije mogao primeniti. Na vrhu su još domeni *number*, *color*, *time_period*, *person* i *quality* koji nisu dalje profinjavani. Domen *doctrines*, kao poddomene, ima *archaeology*, *astrology*, *history linguistics*, *literature*, *philosophy*, *psychology*, *art* i *religion*, i tako dalje. Više o projektu proširivanju Prinstonskog

wordneta semantičkim domenima može se naći u “Integrating Subject Field Codes into WordNet” (Magnini, B. and Cavaglia, G. 2000), “Revising WordNet Domains Hierarchy: Semantics, Coverage, and Balancing” (Bentivogli, L. et. al. 2004) i na stanici wdomains.itc.it.

Drugo proširivanje Prinstonskog wordneta odnosi se na njegovo povezivanje sa takozvanom SUMO ontologijom (Standard Upper Merged Ontology) koju je kao projekat izrade novog standarda otpočeo IEEE 2000. godine. U ovom kontekstu pod ontologijom se smatra rečnik ili glosar čija je struktura takva da omogućava računarsku obradu njegovog sadržaja. Jedna ovakva ontologija se sastoji od koncepata, aksioma i relacija koje opisuju neki domen interesovanja. *Gornja ontologija* ‘Upper Ontology’ ograničava se na meta-koncepte koji su apstraktni ili generički po prirodi i prema tome, dovoljno opšti da mogu da pokriju na višem nivou široki opseg područja. U gornju ontologiju nisu uključeni koncepti koji su specifični za neki određen domen. Termin ‘Merged’ u nazivu ontologije potiče otuda što je ona nastala povezivanjem više javno dostupnih sadržaja u jednu koherentnu strukturu (Pease, A., Niles, I., 2002).

Da bi mogla lako da se razume i primeni, SUMO ontologija se sastoji od relativno malo tvrđenja i koncepata: približno 4.000 tvrđenja, što uključuje preko 800 pravila, i 1.000 koncepata. Neka od opštih tema koje SUMO pokriva jesu:

- strukturni koncepti, kao što su primerci i podklase
- opšti tipovi objekata i procesa
- apstrakcije koje uključuju teoriju skupova, attribute i relacije
- brojevi i mere
- vremenski koncepti, kao što je vremensko trajanje
- delovi i celine
- osnovne semiotičke relacije .

Postavlja se pitanje kako se ontologija može uspešno koristiti u raznovrsnim prirodnojezičkim aplikacijama, kakvo je pronalaženje informacija, koje obrađuju slobodan tekst, to jest, tekst koji nije prethodno obrađen i strukturiran. Jedan odgovor na to pitanje daje povezivanje SUMO ontologije sa velikom leksičkom bazom podataka kakav je Prinstonski wordnet (Niles and Pease 2003). Ovo povezivanje je ostvareno sa verzijom 1.6 Prinstonskog wordneta, ali su ostvarene veze prosleđene i u sve kasnije verzije wordneta. Više o SUMO ontologiji, kao i pregledanje SUMO hijerarhije je obezbeđeno na stranicama sumo.ieee.org i www.ontologyportal.org.

Imajući u vidu činjenicu da se SUMO sastoji od relativno malog broja koncepata, dok je wordnet veoma bogata leksička baza koja je u trenutku povezivanja imala skoro 100.000 sinsetova, potrebno je precizirati na koji način se veza između SUMO koncepata i wordnet sinsetova ostvaruje. U osnovi, korišćene su tri vrste relacija: sinonimija, nadređenost i primerak. Ove vrste relacija biće ilustrovane primerima. U Prinstonskom wordnetu postoji sinset {battle:1, conflict:3, fight:4, engagement:1} (neprijateljski susret suprotstavljenih vojnih snaga u toku rata) koji je sinoniman sa konceptom “Battle” iz SUMO, pa se u sinset dodaje informacija “= Battle”. Ovom sinsetu podređeni sinset je {naval battle:1} (bitka između pomorskih flota) za koji, prirodno, ne postoji sinonimni koncept u SUMO ontologiji. U ovakvom slučaju, sinset se povezuje sa nadređenim konceptom tako da se u ovaj sinset dodaje informacija “+ Battle”. Konačno, sinset {Iwo:1, Iwo Jima:2, invasion of Iwo:1} (krvava i dugačka operacija na ostrvu Iwo Džima u kojoj su se američki marinci iskrcali na ostrvu i porazili japanske branitelje u toku februara i marta 1945. godine) predstavlja jedan primerak, ili slučaj bitke, te se na ovakve sinsetove primenjuje treća vrsta relacije koja ukazuje da koncept označen wordnetom predstavlja jedan član klase koju označava SUMO

koncept. U ovom slučaju sinsetu se dodaje oznaka “@ Battle”. Ne treba da iznenađuje što ima slučajeva kada se više sinsetova iz wordneta povezuje relacijom sinonimije sa istim konceptom iz SUMO. Razlika može da bude lingvistički važna ali sa stanovišta inženjerstva znanja sasvim irelevantna. Tako je informacija “= Battle” pridružena i sinsetovima {invasion:1} (akt kojim jedna armija napada protivničku teritoriju s ciljem da je osvoji ili opljačka) i {combat:1, armed combat:1} (bitka između dve vojne sile).

Ilustracije radi, hijerarhijska grana kojoj pripada ontološki concept „Battle” izgleda ovako:

entity→physical→process→intentional process→
social interaction→contest→violent contest→battle

Sam vrh hijerarhijskog drveta podklasa je:

entity→
physical→
object→
process→
abstract→
quantity→
attribute→
set or class→
relation→
proposition→
graph→
graph element→

Opisano preslikavanje Prinstonskog wordneta u SUMO može da funkcioniše kao prirodnojezički indeks za koncepte iz ontologije, kao most između strukturiranih koncepata iz SUMO ontologije i slobodnog teksta koji je predmet obrade sve većeg broja primena, kao, na primer, konceptualno indeksiranje (Stamou, S. i dr. 2004) ili klasifikacija teksta (Tufiş, D. and Koeva S. 2007).

3 Wordnet i višejezičnost

Kao resurs koji mnogo obećava u prirodnojezičkim aplikacijama, Prinstonski wordnet stiže veliku popularnost koja potiče od toga što je podstakao projekte izgradnje sličnih resursa i za druge jezike. Jedan od prvih značajnih projekata bio je projekat EuroWordNet koji je od 1996. do 1999. godine finansirala Evropska zajednica u okviru programa FW4. Cilj ovog projekta bila je izgradnja višejezičke leksičke baze podataka koja bi sadržala wordnete za osam evropskih jezika: engleski, holandski, italijanski, španski, francuski, nemački, češki i estonski. Struktura wordneta svih ovih jezika odgovara strukturi Prinstonskog wordneta, ali je svaki od njih autonoman i sadrži skup koncepata koji odgovara specifičnostima leksikalizacije u svakom pojedinačnom jeziku. Međutim, da bi se zadovoljile potrebe višejezičnih primena čiji značaj u kontekstu Interneta i veća postaje sve veći, u okviru projekta EuroWordNet uveden je pojam interlingvalnog indeksa – skraćeno ILI – preko koga se sinset jednog jezika povezuje sa sinsetovima u drugim jezicima koji predstavljaju slične koncepte (Vossen, P. 2004). Svrha interlingvalnog indeksa je da omogući efikasno preslikavanje između inače autonomnih struktura pojedinačnih jezika. Pošto svaki jednojezički wordnet predstavlja strukturu za sebe, sam ILI se svodi na kondenzovani univerzalni indeks značenja. S obzirom na nezavisnost pojedinačnih wordneta i pretpostavljenu različitost leksikalizacije koncepata u različitim jezicima sinsetovi se povezuju sa ILI indeksom različitim relacijama koje su u

principu “mnogo-mnogo”. Osim povezivanja sinsetova jednog jezika sa konceptulano sličnim sinsetovima drugih jezika, ovakva višezjezička baza podataka omogućava da se deli znanje koje je jezički nezavisno. Tako ILI indeks omogućava da se dodatno znanje koje je uvedeno u Prinstonski wordnet, a o kome je bilo govora u prethodnom odeljku, obeležje domena i SUMO koncepta, može koristiti i wordnetima drugih jezika.

Projekat BalkaNet koji je od 2001. do 2004. godine finansirala Evropska komisija (Tufiş, D. et al. 2004) direktno se nastavlja na dostignuća projekta EuroWordNet i dalje ih produbljuje. Cilj ovog projekta bio je razvoj poravnatih mreža tipa wordnet za balkanske jezike, i to bugarski, grčki, rumunski, srpski i turski, kao i proširenje mreže za češki koja je početno bila razvijana u okviru projekta EuroWordnet. Osnovni cilj BalkaNet projekta bio je razvoj savremenih jezičkih resursa za balkanske jezike koji bi omogućili nov način pristupa informacijama koje potiču iz balkanskih jezika. Osim toga, cilj ovog projekta bio je i proširivanje višezjezičke baze koja je uspostavljena u okviru projekta EuroWordnet balkanskim jezicima.

Kao glavne aktivnosti u okviru Balkanet-a treba istaći, pre svega, razvoj mreža wordnet za balkanske jezike pojedinačno i njihovo povezivanje sa postojećom leksičkom bazom EuroWordNet. Ove glavne aktivnosti su planirane i sprovedene sinhronizovano, što znači da su jednojezičke mreže izgrađene nad zajednički dogovorenim osnovnim skupovima koji su već bili prisutni u Prinstonskom wordnetu. Izvan ovih osnovnih skupova, za svaki pojedinačni jezik mreža se razvijala nezavisno, ali uvek u okvirima koje postavlja Prinstonski wordnet. Ovakav pristup razvoju wordneta postavio je specifične probleme. Naime, tokom rada na razvoju mreže često su se postavljala sledeća pitanja: da li su koncepti jezički zavisni ili ne, da li su obrasci za leksikalizaciju koncepta univerzalni, da li je struktura Prinstonske mreže valjana i za druge jezike, i da li je skup semantičkih relacija koje su u njega ugrađene dovoljan za sve jezike (Vossen, P. 2004). Premda je rad na razvoju zasebnih mreža za balkanske jezike često davao potvrde za negativan odgovor na ova pitanja, nije se odustalo od prethodno utvrđenog postupka. Kako se mreže wordnet danas razvijaju pre svega za informatičke potrebe, tako se i osnovna primena ovih mreža za balkanske jezike vidi u njihovoj ugradnji u informatičke primene zasnovane na prirodnojezičkoj obradi, na primer, za klasifikaciju dokumenata na mreži i višezjezičko pretraživanje. Postojanje višezjezičke baze sa međusobno poravnatim konceptima je tada od suštinskog značaja.

The image shows three side-by-side windows of WordNet interfaces. The left window is for English WordNet 2.0.c (t), the middle for Serbian WordNet 0.4.a, and the right for Bulgarian WordNet 1.0.a (t). Each window displays the entry for the word 'can' and its hierarchical structure. The English entry shows 'tin can' with various senses and relationships. The Serbian entry shows 'konzerva' with its own set of relationships. The Bulgarian entry shows 'консервна кутия' with its own set of relationships. The interfaces include tabs for 'View', 'Tree', 'RevTree', and 'Edit', as well as buttons for 'BCS1,2', 'BCS3', and 'XML'.

Na slici je prikazana leksikalizacija pojma “hermetički zatvoren metalni sud za hranu, piće ili boju itd.” u engleskom ({can:1, tin:3, tin can:2}), srpskom ({konzerva:1}) i bugarskom ({консервна кутия:1}). Dok leksikalizacija ovog pojma ni u jednom od jezika nije sporna, njihovi nadređeni pojmovi koji su leksikalizovani u engleskom kao {container:1}, {instrumentality:1, instrumentation:1} i {artifact:1, artefact:1} su prilično veštački preslikani iz Prinstonskog wordneta i u srpski i u bugarski wordnet.

4. Srpski wordnet i njegova izgradnja

Izgradnja Srpskog wordneta otpočela je sa Balkanet projektom, i po završetku projekta baza podataka sadržala je 8059 sinsetova od koji je 7736 nastalo preuzivanjem iz Prinstonskog wordneta, dok je 117 pripadalo takozvanom skupu balkanskih specifičnih koncepata, a 206 skupu srpskih specifičnih koncepata. Jedan koncept specifičan za Balkan koji je poznat i leksikalizovan su svim balkanskim jezicima, a koga nema u Prinstonskom wordnetu je {alva:1} u sprskom, {халва:1} u bugarskom, {χαλβάς:1} u grčkom, {halva:1} u rumunskom i {kağıt helva:1} u turskom. Po završetku projekta rad na sprskom wordnetu je nastavljen ali ne u okviru nekog formalnog projekta, već se više zasnivao na volonterskom radu. U prvoj godini posle završetka projekta najviše je urađeno na izradi sinsetova iz domena bilogije koji obuhvata biljne i životinjske vrste, kao i više klasifikacione grupe kojima te vrste pripadaju. Izbor domena bio je usklađen sa dogradnjom srpskog morfološkog elektronskog rečnika odrednicama iz istog domena. U ovom periodu dosta je urađeno i na dopuni Srpskog wordneta balkanski specifičnim i sprski specifičnim konceptima (Krstev, C. 2006).

Od početka 2006. godine otpočeo je kooperativan rad na daljoj dogradnji Srpskog wordneta. Naime, postdiplomske studije na Grupi za bibliotekarstvi i informatiku Filološkog fakulteta Univerziteta u Beogradu upisuju svršeni studenti različitih profila od kojih veliki broj radi kao bibliotekar, informatičar ili dokumentalista u javnim ili specijalnim bibliotekama. U okviru obaveznog predmeta na postdiplomskim studijama studenti treba da urade jedan seminarski rad. Javila se ideja da bi ovi studenti s obzirom na prirodu svog posla mogli da urade jedan segment wordneta koji bi odgovarao njihovom specifičnom znanju koje su stekli u toku prethodnog školovanja. Ovom poduhvatu za koji su uključeni studenti pokazali veliki entuzijazam, priključili si se i volonteri sa drugih studijskih grupa. U narednom odeljku biće opisan njihov rad.

Za odabir podskupa sinsetova koji najviše odgovaraju pojedinačnim studentima korišćen je softverski alat WS4LR (*Work Station for Lexical Resources*) koji je detaljnije opisan u radu (Obradović, I. i Stanković, R. 2008). S obzirom da Srpski i Prinstonski wordnet za potrebe prenosa i razmene koriste XML format koji je uveden u okviru Balkanet projekta, ovaj alat omogućava korisniku da sam formuliše *XML Path* izraz kojim može da selektuje sinsetove iz izabranog wordneta. Za ovakvu selekciju korišćena je najčešće kombinovana informacija o pripadnosti domenu i ontološkoj kategoriji jer sama informacija o domenu često proizvodi suviše veliki podskup sinsetova. Tako, na primer, Prinstonski wordnet ima 1181 sinset iz domena prava, što bi bio suviše veliki zadatak za seminarski rad jednog studenta. S druge strane, ako bi se koristile samo ontološke kategorije mogli bi se dobiti pojmovi koji pripadaju različitim domenima za koje izabrani student nije kompetentan. Na primer, ontološka kategorija "Charcater" osim domenu "linguistics" za koji smo bili zainteresovani, pripada i domenima "factotum", "number", "publishing", i tako dalje. Takođe, da bi izabrano područje bilo iscrpno obrađeno ponekada je trebalo dopuniti selektovani podskup. Na primer, jedan selektovani podskup iz domena lingvistike odnosio se na ontološku kategoriju "NaturalLanguage". Obrada i uključivanje ovog iscrpnog podskupa u srpski wordnet je pokazala da i dalje nisu uključeni neki od "velikih" evropskih jezika, kao što su ruski, francuski, itd. Pokazalo se da su oni povezani sa zasebnim ontološkim kategorijama "=RussianLanguage" i "=FrenchLanguage". Ovi nedostajući sinsetovi su kasnije uključeni u Srpski wordnet.

4.1. Domen lingvistike – Bojana Đorđević

U projekat izrade srpskog wordneta uključila sam se početkom 2007. godine, nakon diplomiranja na katedri za Opštu lingvistiku na Filološkom fakultetu u Beogradu. Moj zadatak na ovom projektu bio je preuzimanje sinsetova iz domena lingvistike iz Prinstonskog

wordneta i njihovo prilagođavanje za Srpski wordnet. Obradeni skup sinsetova iz domena lingvistike obuhvatao je sledeće ontološke kategorije: morfeme (16 sinsetova), gramatiku (238), karaktere (87) i prirodne jezike (595). Imajući u vidu sinsetove koji su naknadno dodati skupu radi obezbeđivanja pravilnog hijerarhijskog povezivanja sa preostalim delom Srpskog wordneta, iz domena lingvistike prilagođeno je ukupno 946 sinsetova.

Tipičan primer jednog prevedenog sinseta, preuzetog is skupa prirodnih jezika, izgleda ovako:

<pre><SYNSET> <ID>ENG20-06480396-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>mother tongue <SENSE>1</SENSE></LITERAL> <LITERAL>maternal language <SENSE>1</SENSE></LITERAL> <LITERAL>first language <SENSE>1</SENSE></LITERAL> </SYNONYM> <ILR> <TYPE>hypernym</TYPE> ENG20-06479855-n</ILR> <DEF>one's native language; the language learned by children and passed from one generation to the next</DEF> <DOMAIN>linguistics</DOMAIN> <SUMO>NaturalLanguage<TYPE>+</TYPE></SUMO> </SYNSET></pre>	<pre><SYNSET> <ID>ENG20-06480396-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>maternji jezik <SENSE>1</SENSE></LITERAL> <LITERAL>prvi jezik <SENSE>1</SENSE></LITERAL> <LITERAL>rođeni jezik <SENSE>1</SENSE></LITERAL>¹ </SYNONYM> <ILR> <TYPE>hypernym</TYPE> ENG20-06479855-n</ILR> <DEF>jezik koji je najpre usvojen u detinjstvu ili onaj kome se daje prednost u višejezičnoj situaciji</DEF> <SNOTE>Uradila B. Đorđević, postdiplomac C. Krstev</SNOTE> </SYNSET></pre>
---	---

U toku rada na Srpskom wordnetu, nailazila sam na nekoliko tipova problema. S obzirom na to da su nazivi jezika sačinjavali veliki broj sinsetova koje je trebalo uključiti u Srpski wordnet, osnovni problem je bio pronaći adekvatne i, po mogućstvu, odomaćene nazive za pojedine manje istražene, a u engleskom wordnetu precizno urađene jezike. Ovo se, pre svega, odnosilo na amerindijanske jezike, ali i na grupe jezika poput afričke.

Druga vrsta problema nastajala je prilikom pokušaja da se pronađe odgovarajući termin za pojavu koja ili ne postoji u srpskom jeziku ili je klasifikovana na drugi način. Navedeni primer odnosi se na vrstu objekta kakva ne postoji u srpskom. Autor upotrebljenog termina je Ljiljana Mihailović (Mihailović, Lj.1967):

<pre><SYNSET> <ID>ENG20-05923070-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>retained object <SENSE>1</SENSE></LITERAL> </SYNONYM> <ILR> <TYPE>hypernym</TYPE>ENG20-05922459- n</ILR> <DEF>an object in a passive construction</DEF> <DOMAIN>grammar</DOMAIN> <SUMO>NounPhrase<TYPE>+</TYPE></SUMO> </SYNSET></pre>	<pre><SYNSET> <ID>ENG20-05923070-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>zadržan objekat <SENSE>1</SENSE></LITERAL> </SYNONYM> <ILR> <TYPE>hypernym</TYPE> ENG20-05922459-n</ILR> <DEF>objekat u pasivnoj konstrukciji</DEF> <SNOTE>Uradila B. Đorđević, postdiplomac C. Krstev</SNOTE> </SYNSET></pre>
--	--

Slično tome, neke pojave koje se u engleskom mogu opisati jednom rečju (*to punctuate*), u srpskom su se morale prevesti opisno (*obeležiti znacima interpunkcije*).

¹ Termini “prvi jezik” i “maternji jezik” preuzeti su iz *Kembričke enciklopedije jezika*, (Kristal, 1995),

Najveći broj termina i definicija preuzet je iz *Kembričke enciklopedije jezika* (Kristal, 1995), *Enciklopedijskog rečnika moderne lingvistike* (Kristal, 1998), *Uvoda u opštu lingvistiku* (Bugarski, 1996) i *Gramatike engleskog jezika* (Đorđević, 2002). U prevođenju naziva jezika, konsultovane su studije *Jeziči* (Bugarski, 1996a) i *Jezik* (Sapir, 1992), koje su bile od posebne pomoći pri prevođenju naziva amerindijanskih jezika. Po potrebi su korišćeni i svi dostupni rečnici: engleskog jezika (Institut za strane jezike, 2005), stranih reči (Klajn, Šipka, 2006) i sinonima (Ćosić, 2007).

Internet baze podataka bile su od neprocenjivog značaja za prevođenje naziva jezika. Jedan od glavnih izvora bila je *Vikipedija*, kako na srpskom, tako i na hrvatskom jeziku. Tu je zatim i popis osnovnih jezičkih grupa i podgrupa Danka Šipke, deo opštije tematske liste namenjene studentima srpskog kao stranog jezika. Treba pomenuti i hrvatsku bazu *PhraseBASE* u kojoj je po jezičkim porodicama precizno grupisan velik broj jezika, kao i tekst Nevenke Hajdarović *Izmjene i dopune u UDK i njegova primjena u BH bibliotekarstvu*, koji sadrži opširan popis svetskih jezika. Koristila sam i bogatu bazu naziva jezika na koje automatski prevodilac PROZ prevodi sa srpskog (i obrnuto).

Kada bi se desilo da se, poređenjem izvora sa Interneta, dobije nerealno veliki broj mogućih prevoda, učestalost u rezultatima pretrage u Guglu imala je konačnu reč. Ipak, učestalost ili pojavljivanje i nepojavljivanje među rezultatima u nekim situacijama nisu mogli biti od presudnog značaja zbog još uvek nedovoljnog broja online lingvističkih tekstova na srpskom jeziku.

Što se konsultanata tiče, u nekoliko navrata dobila sam korisne informacije od članova liste slanja *ST-L*, koju je pokrenuo je Prof. Dr Danko Šipka kao početni korak u planiranju srpskog elektronskog korpusa. Lista je najpre bila namenjena diskutovanju o eventualnom sastavu korpusa srpskog (kao i bosanskog i hrvatskog) jezika i formi tekstova koji bi u njega ušli, a kasnije dobila i opštelingvistički karakter. Posebno se zahvaljujem dvojici članova: Wales Brown-u, profesoru lingvistike sa Kornel univerziteta, SAD, i Pavlu Ćosiću, lingvisti i autoru Rečnika sinonima, na opštejezičkim savetima.

4.2. Domen biomedicine – Sanja Antonić

Praktičan rad na razvoju srpskog wordneta predstavlja i poslednje poglavlje moje magistarske teze pod nazivom „Razvoj informatičke semantičke mreže za oblast biomedicine“ koja je rađena pod mentorstvom profesorke Cvetane Krstev. Zaposlena sam u Univerzitetnoj biblioteci „Svetozar Marković“ u Beogradu, Odeljenje za naučne informacije kao viši stručni saradnik za biomedicinu i biotehnologiju. Diplomirala sam molekularni biologiju i fiziologiju, a stečeno znanje sam primenjivala i proširivala dodajući nove koncepte za oblast biomedicine u okviru srpskog wordneta.

Biomedicina obuhvata veliki broj naučnih disciplina a većina njih su vrlo dinamične i veoma se brzo razvijaju. Problem zastarevanja teorije ili terminologije u biomedicini je veoma prisutan, tako da je potrebno pratiti nova saznanja i otkrića gotovo svakodnevno. Usled toga se mnogi naučni i stručni termini veoma često i ne prevode sa engleskog jezika. U toku rada trebalo je voditi računa i o ciljnoj populaciji kojoj je mreža namenjena i ponekad je bilo teško uskladiti potrebe stručnjaka za obrađivane oblasti ali i šireg kruga korisnika. Posebno sam vodila računa da mreža treba da bude primenjiva u informatičkim aplikacijama.

Tokom rada na srpskom wordnetu bavila sam se sledećim naučnim disciplinama koje su posedovale svoje specifične osobine kao i terminologiju: citologija, histologija, embriologija, genetika, virusologija odnosno mikrobiologija, zoologija beskičmenjaka i kičmenjaka, veterina, poljoprivreda itd. Konkretno, bavila sam se prilagođavanjem onih delova Prinstonskog wordneta (PWN) za srpski jezik koji pripadaju domenu biologije, a prema SUMO ontologiji povezani su sa sledećim ontološkim kategorijama: Cell – ćelija,

Genetics – genetika, Virus – virusi, Bacterium - bakterije, Microorganism – mikroorganizmi, ScienceFields – naučne oblasti, ukupno 462 sinseta.

Tipičan primer jednog prevedenog sinseta koji pripada ontološkoj kategoriji “Mikroorganizmi” je prikazan na sledeći način:

<pre> <SYNSET> <ID>ENG20-01298897-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>mycoplasma <SENSE>1</SENSE> </LITERAL> </SYNONYM> <ILR>ENG20-01280902-n <TYPE>hypernym</TYPE> </ILR> <ILR>ENG20-01298746-n <TYPE>holo_member</TYPE> </ILR> <DEF>the smallest self-reproducing prokaryote; lacks a cell wall and can survive without oxygen; can cause pneumonia and urinary tract infection</DEF> <DOMAIN>biology</DOMAIN> <SUMO>Bacterium <TYPE>+</TYPE> </SUMO> <RILR>ENG20-01299130-n <TYPE>hypernym</TYPE> </RILR> </SYNSET> </pre>	<pre> <SYNSET> <ID>ENG20-01298897-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>mikoplazma <SENSE>1</SENSE> </LITERAL> </SYNONYM> <ILR>ENG20-01280902-n <TYPE>hypernym</TYPE> </ILR> <ILR>ENG20-01298746-n <TYPE>holo_member</TYPE> </ILR> <DEF>najmanje prokariote koje mogu da se bespolno razmnožavaju; nedostaje im ćelijski zid i mogu preživeti bez kiseonika; mogu izazvati pneumoniju (upalu pluća) i infekcije urinarnog trakta</DEF> <SNOTE>Uradila S. Antonić, postdiplomac C. Krstev</SNOTE> <RILR>ENG20-01299130-n <TYPE>hypernym</TYPE> </RILR> </SYNSET> </pre>
--	---

Mikoplazme su karakteristični mikroorganizmi. Doslovan prevod glose iz PWN koja glasi „the smallest self-reproducing prokaryote“ bio bi „najmanje prokariote koje se samoreprodukuju“, ali poznavanje životnog ciklusa organizama, pokazuje da je stručniji i samim tim bolji prevod „najmanje prokariote koje se bespolno razmnožavaju“. Ovaj primer pokazuje da ukoliko želimo da proizvedemo kvalitetnu mrežu wordnet, ni prevod glose ne može se automatski uraditi.

Sinset {paramecium:1, paramecia:1} koji reprezentuje organizme iz roda Paramecium predstavljaju redak primer koncepta za koji u srpskom jeziku pored stručnog termina odnosno njegovog latinskog naziva *paramecium* postoji i termin na srpskom jeziku *papučica*, koji je u širokoj upotrebi i nalazimo ga u mnogim udžbenicima za osnovnu i srednju školu.

Mnogo je češći slučaj da u Prinstonskom wordnetu nalazimo više literala za određeni pojam i samo jedan ili dva adekvatna na srpskom jeziku. Ilustrativan je primer prideva *amebni*, za koji postoji čak pet sinonima na engleskom jeziku {amoebic:1, amebic:1, ameban:1, amoeban:1, amoebous:1} koji su u suštini svi varijantni oblici.

Praktičan rad na izradi srpskog wordneta je u većini slučajeva bio pravi istraživački zadatak, pri čemu smo koristili klasične izvore, prvenstveno štampane rečnike i udžbenike, a zatim bi usledila provera na Internetu, korišćenjem Google-a ili proverenih akademskih, stručnih i obrazovnih veb stanica. Ovaj istraživački rad se najbolje može ilustrovati primerom sinseta koji se odnosi na veoma dobro izučenu naslednu bolest, hemofiliju. Postoje različite vrste hemofilije u zavisnosti od vrste genetskog poremećaja faktora koagulacije koji je izazivaju. Jedna od njih je *hemofilija B*, za koju se na engleskom jeziku koristi i termin *Christmas disease*. U prvom trenutku je izgledalo da za taj termin u srpskom ne postoji odgovarajući ili da bi ga mogli prevesti sa *Božićna bolest*, s obzirom da *Christmas* na engleskom znači *Božić*. Odgovor na ovu nedoumicu je nađen u klasičnom udžbeniku, Kičićevoj knjizi „Medicinska genetika“ (Kičić & Krajičanić, 1989) koji govori o

Kristmasovoj bolesti, tako da konačno sinsetovi koji odgovaraju ovoj bolesti u PWN i srpskom wordnetu izgledaju ovako:

<pre> <SYNSET> <ID>ENG20-13364794-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>hemophilia B <SENSE>1</SENSE> </LITERAL> <LITERAL>haemophilia B <SENSE>1</SENSE> </LITERAL> <LITERAL>Christmas disease <SENSE>1</SENSE> </LITERAL> </SYNONYM> <ILR>ENG20-13364162-n <TYPE>hypernym</TYPE> </ILR> <DEF>a clotting disorder similar to hemophilia A but caused by a congenital deficiency of factor IX</DEF> <DOMAIN>genetics</DOMAIN> <SUMO>DiseaseOrSyndrome <TYPE>+</TYPE> </SUMO> </SYNSET> </pre>	<pre> <SYNSET> <ID>ENG20-13364794-n</ID> <SYNONYM>- <LITERAL>hemofilija B <SENSE>1</SENSE> </LITERAL> <LITERAL>Kristmasova bolest <SENSE>1</SENSE> </LITERAL> </SYNONYM> <DEF>poremećaj zgrušavanja veoma sličan hemofiliji A ali se javlja usled kongenitalne deficijencije faktora IX (po imenu dečaka Christmas, prvi put dijagnostikovana 1952. godine)</DEF> <SNOTE>Uradila S.Antonić, postdiplomac C. Krstev</SNOTE> <POS>n</POS> <ILR>ENG20-13364162-n <TYPE>hypernym</TYPE> </ILR> </SYNSET> </pre>
---	---

Tokom praktičnog rada koristila sam veći broj rečnika: „Englesko-srpski rečnik“ (Benson, M. 1997), „The new Merriam-Webster dictionary“ (Merriam-Webster 1989), rečnik srpskohrvatskog književnog jezika (Matica srpska i Matica hrvatska, 1967) i „Veliki rečnik stranih reči i izraza“ (Klajn, I., Šipka, M. 2006).

S obzirom da opšti rečnici uglavnom ne sadrže stručne pojmove iz oblasti koje sam obrađivala, u radu su korišćeni i mnogi udžbenici. Za oblast mikrobiologije korišćeni su udžbenici (Jemcev, V. T. i Đukić, D. 2000), (Tešić, Ž. i Todorović, M. 1992), (Jarak, M. i Govedarica, M. 2003), a za oblast genetike (Marinković, D. i sar., 1989), (Kičić, M. i Krajičanić, B. 1989) i (Dumanović, J. i sar., 1985). Osim toga, korišćeni su i sledeći priručnici: *Zoologija Invertebrata* (Krunić, M. 1990), *Citologija* (Grozdanović-Radovanović, J. 1985), *Histologija* (Grozdanović-Radovanović, J. 1980), *Razviće životinja* (Ćurčić, B. 1985). Veoma interesantan, ali i zahtevan je bio rad na sinsetovima iz oblasti virusa jer oni spadaju u najsitnije i po građi najjednostavnije mikroorganizme koji imaju veoma visoku stopu mutacija i zbog toga su veoma teški za izučavanje. Od neprocenjive pomoći bio je sjajan udžbenik profesora Ljubiše Krstića *Medicinska virusologija* (Krstić, Lj. 2005) kao najpouzdaniji izvor za terminologiju vezanu za viruse.

Neočekivano korisna rešenja za obradu sinsetova koji pokrivaju naučne oblasti iz domena biologije našla sam na Internet stranicama Vikipedija i veb stanici Vokabular. Ostali korisni Internet izvori bili su Rečnik Filozofskog fakulteta u Novom Sadu koji sam koristila tokom 2006. godine ali koji više nije aktivan, Zaštita bilja i Human Genome Project Information.

4.3. Domen religija – Nevena Ivković-Berček

S obzirom da sam diplomirani teolog i bibliotekar u biblioteci Teološkog fakulteta u Beogradu termini vezani za religiju i verska literatura su mi bili relativno dobro pozanti jer su bili sastavni deo mojih studija a sa njima se srećem i u okviru mog sadašnjeg posla.

Moj zadatak u ovom projektu bio je preuzimanje iz Prinstonskog wordneta sinsetova iz domena religije i njihovo prilagođavanje za Srpski wordnet. Obradeni skup sinsetova obuhvataju ontološku kategoriju ‘verski proces’ (Religious Process) (130 sinsetova) i ‘verske organizacije’ (Religious Organizations) (160 sinsetova).

Tipičan primer jednog prevedenog sinseta koji pripada ontološkoj kategoriji ‘verski proces’ izgleda ovako:

<pre><SYNSET> <ID>ENG20-00979294-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>anointing of the sick <SENSE>1</SENSE> </LITERAL> <LITERAL>extreme unction <SENSE>1</SENSE> </LITERAL> <LITERAL>last rites <SENSE>1</SENSE> </LITERAL> </SYNONYM> <ILR>ENG20-00974693-n <TYPE>hypernym</TYPE> </ILR> <DEF>a Catholic sacrament; a priest anoints a dying person with oil and prays for salvation</DEF> <DOMAIN>religion</DOMAIN> <SUMO>ReligiousProcess <TYPE>+</TYPE> </SUMO> </SYNSET></pre>	<pre><SYNSET> <ID>ENG20-00979294-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>poslednja pričest <SENSE>1</SENSE> </LITERAL> </SYNONYM> <ILR>ENG20-00974693-n <TYPE>hypernym</TYPE> </ILR> <DEF>katolička sveta tajna; sveštenik miropomazuje umiruću osobu uljem i moli se za spasenje</DEF> <SNOTE>Uradila N. Ivković, postdiplomac C. Krstev</SNOTE> </SYNSET></pre>
--	---

Pošto sam se u radu bavila terminima koji su vezani za verske organizacije i verske procese, susrela sam se sa dve vrste problema. Poznato je da u srpskom jeziku nedostaju mnogi termini za strane reči. Sa tim problemom sam se susrela u toku pevođenja naziva verskih procesa koji su našoj kulturi nepoznati i nekih verskih zajednica koje ne postoje na našim prostorima. Na primer, s oznakom ‘verski proces’ u Prinstonski wordnet je uključen sinset {popery:1} čije značenje je ponižavajući ili uvredljiv izraz za običaje i rituale Rimske katoličke crkve. U našem jeziku ovaj koncept nije poznat niti se koristi.

Među verskim organizacijama u Prinstonskom wordnetu našla se protestantska denominacija osnovana 1886. godine od strane Meri Bejker Edi čiji naziv je *Christian Science* ili *Church of Christ Scientist*. Na našim prostorima ne postoji ovaj pokret pa za nju ni ne postoji naziv koji bih mogla da preuzmem. Zato sam se uz konsultacije sa prof. Krstev opredelila za naziv “Crkva Hristovih naučnika”.

Od stručnih leksikona, rečnika i enciklopedija koristila sam studije *Udar na verske slobode* (Bjelajac, B. i Vidović, D. 2001), *Sekte i politika* (Branković, T. 2000) i *verske sekte* (Đurđević-Stojković, B. 2002), *Enciklopediju živih religija* (Krim, K. 1992) i *Rječnik biblijske teologije* (Leon-Dufour, X. 1969). Kako stručna literatura često nije posedovala prevod ili objašnjenje termina na srpskom jeziku, morala sam da koristim opšte rečnike stranih reči (Anić, Š. i Klaić, N. 2002; Vujaklija, M. 2005; Klaić, B.1951; Klajn, I. i Šipka, M. 2006) i rečnike engleskog jezika (MacMillan, 2002), srpskog jezika (Moskovljević, M. 1966) i dvojezične engleskosrpske rečnike (Ristić, S. i dr., 1956).

Internet izvori koje sam najčešće koristila i koji su mi dosta pomogli u radu jesu: Miriam Webster, srpska i engleska Wikipedia, Vokabular, koji je zamišljen kao slobodan servis dostupan svim zainteresovanima, nudeći im opšti srpsko-srpski rečnik i druge servise

koji iz rečnika mogu da izniknu, i Metak kao zaseban rečnički servis stanice SerbianCafe, a pored njih sam koristila i različite pretraživače, ali sam do potrebnih informacija u većini slučajeva dolazila putem Gugla i Krstarice.

4.4. Domen literature – Zorica Zorica

Osnovne studije završila sam na Filozofskom fakultetu u Novom Sadu na katedri za srpsku književnost i jezik, a sada radim kao bibliotekar u Galeriji likovne umetnosti poklon zbirka Rajka Mamuzića u Novom Sadu.

Moj zadatak u ovom projektu bio je preuzimanje iz Pristonskog wordneta sinsetova iz domena književnosti i njihovo prilagođavanje za srpski Wordnet; to su dakle sinsetovi koji u PWN pripadaju domenu «literature». Najveći broj obrađenih sinsetova pripada ontološkim kategorijama "tekst" (Text) (218), "pisanje" (Writing) (17), "lingvistički izraz" (Linguistic expression) (16), "razvoj sadržaja" (ContentDevelopment) (11) i drugih kategorija iz domena teorije književnosti i retorike. Ukupan broj obrađenih sinsetova je 355. Tipičan primer jednog prevedenog sinseta koji pripada ontološkoj kategoriji "tekst" izgleda ovako:

<pre> <SYNSET> <ID>ENG20-05976529-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>fable <SENSE>2</SENSE></LITERAL> <LITERAL>parable <SENSE>1</SENSE></LITERAL> <LITERAL>allegory <SENSE>1</SENSE></LITERAL> <LITERAL>apologue <SENSE>1</SENSE></LITERAL> </SYNONYM> <ILR> <TYPE>hypernym</TYPE> ENG20-05974336-n</ILR> <ILR> <DEF>a short moral story (often with animal characters)</DEF> <DOMAIN>literature</DOMAIN> <SUMO>Text<TYPE>+</TYPE></SUMO> </SYNSET> </pre>	<pre> <SYNSET> <ID>ENG20-05976529-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>basna <SENSE>2</SENSE></LITERAL> <LITERAL>parabola <SENSE>1</SENSE></LITERAL> <LITERAL>alegorija <SENSE>1</SENSE></LITERAL> <LITERAL>apolog <SENSE>1</SENSE></LITERAL> <LITERAL>poučna priča <SENSE>1</SENSE></LITERAL> <LITERAL>poučna basna <SENSE>1</SENSE></LITERAL> </SYNONYM> <ILR> <TYPE>hypernym</TYPE> ENG20-05974336-n</ILR> DEF>kratka poučna priča u kojoj su često junaci životinje</DEF> <SNOTE>Uradila Z. Zorica, postdiplomac C. Krstev</SNOTE> </SYNSET> </pre>
--	--

U toku rada na izradi srpskog Wordneta susretala sam se sa nekoliko vrsta problema. Jedan od glavnih problema je bio kako ustanoviti šta je adekvatan prevod koji odgovara duhu našeg jezika, a koji nije neologizam.

Druga vrsta problema nastupala je pri pokušaju da pronađem odgovarajući termin za koncepte koji u našem jeziku nisu leksikalizovani ili su iskazani nekim zajedničkim terminom. Neke termine kao što su «sloganeer» (praviti novo geslo) ili «novelization», «novelisation» (prebacivanje nečeg u formu romana), ostavila sam neprevedene. I termin «potboiler» (književna kompozicija sa slabim kvalitetom koja je napisana da bi se brzo zaradio novac) sam takođe ostavila nepreveden jer u našoj književnoj terminologiji ne postoji precizan termin za ovaj pojam. Najpribližniji su mu pojmovi trivijalna književnost ili petparački roman koji ipak ne označavaju isti koncept npr. potboiler ne mora da bude roman dok trivijalna književnost nije nužno pisana radi brzog zarađivanja.

Treća vrsta problema bili su termini kao što su "novelte", "novella" koji u anglosaksonskoj književnoj terminologiji imaju značenje kratkog romana, dok u srpskoj književnoj terminologiji imaju značenje "vrlo kratke novele" i "vrste pripovetke" te sam te termine prevela sa značenjem koje imaju u anglosaksonskoj terminologiji, mada bi trebalo staviti neku sugestiju sa našim značenjem. Takođe i termin "romance" u značenju srednjovekovne priče u stihu ili prozi ostavila sam u izvornom obliku jer je u tom obliku poznat u našoj književnoj terminologiji. Ipak, treba imati u vidu da se kod nas pod terminom "romansa" pre svega misli na "epsko-lirsku narodnu pesmu". Slično je i sa terminom "story" (priča) za koji se u srpskoj književnoj terminologiji češće upotrebljava termin "pripovetka" kao prozna vrsta.

Pri izradi sinsetova iz domena književnosti za Srpski wordnet koristila sam više rečnika i leksikona: *Rečnik književnih termina* (Živković, D. 1992), *Leksikon stranih reči i izraza* (Vujaklija, M. 1986), *Veliki rečnik stranih reči i izraza* (Klajn, I. i Šipka, M. 2007.), *Bibliotekarski terminološki rečnik* (Kovačević, Lj. 2004), i udžbenik *Teorija književnosti sa teorijom pismnosti* (Živković, D. 2001).

Od internet izvora podataka koristila sam najčešće Vokabular.org, Rastko.org i pretaživače Google, Yahoo i Krstaricu koji su se pokazali kao šturi po pitanju srpskih književnih termina.

4.5. Domen prava – Vesna Crnogorac

Kao diplomiranim pravniku termini koji se odnose na pravo u najširem smislu reči bili su mi najpogodniji i prihvatljivi, budući da sam se sa njima najviše sretala i na osnovnim studijama pravnih nauka i kasnije u petogodišnjem radu na poslovima pravника. Sledeće četiri godine sam radila kao novinar, dok poslednjih deset godina radim kao bibliotekar na različitim poslovima. Od 2006. godine sam sekretar Bibliotekarskog društva Srbije na profesionalnom radu.

Moj zadatak u ovom projektu bio je preuzimanje iz Prinstonskog wordneta sinsetova iz domena prava i njihovo prilagođavanje za Srpski wordnet. Pravni termini su se odnosili na većinu važećih grana prava u našem pravnom sistemu: krivično i krivično – procesno pravo; obligaciono pravo, trgovinsko i međunarodno trgovinsko pravo, građansko i građansko procesno pravo, upravno pravo; međunarodno – javno pravo; nasledno pravo; penologija; do osnovnih pojmova koji se obrađuju u okviru predmeta - uvod u pravo. U Prinstonskom wordnetu ima mnogo sinsetova koji pripadaju domenu 'law', a kako nisu svi mogli da budu obrađeni ovom prilikom kriterijum za odabir je bio sledeći: (a) da pripadaju trećem skupu baznih koncepata koji je ustanovljen u toku projekta Balkanet, a da još uvek nisu uključeni u srpski wordnet (ukupno 42 sinseta); (b) da pripadaju ontološkoj kategoriji 'Certificate' (ukupno 73 sinseta). Tako je ukupno obrađeno 115 sinsetova.

Tipičan primer jednog prevedenog sinseta koji pripada krivično procesnom pravu izgleda ovako:

<pre> <SYNSET> <ID>ENG20-01122850-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>conviction <SENSE>2</SENSE></LITERAL> <LITERAL>judgment of conviction <SENSE>1</SENSE></LITERAL> <LITERAL>condemnation <SENSE>5</SENSE></LITERAL> <LITERAL>sentence <SENSE>2</SENSE></LITERAL> </pre>	<pre> <SYNSET> <ID>ENG20-01122850-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>osudujuća presuda <SENSE>5</SENSE></LITERAL> </pre>
--	--

<pre> </SYNONYM> <ILR><TYPE>hypernym</TYPE> ENG20-01122569-n</ILR> <ILR><TYPE>near_antonym</TYPE> ENG20-01127432-n</ILR> <ILR><TYPE>eng_derivative</TYPE> ENG20-00876567-v</ILR> <ILR><TYPE>eng_derivative</TYPE> ENG20-00876935-v</ILR> <ILR><TYPE>category_domain</TYPE> ENG20-06135956-n</ILR> <DEF>(criminal law) a final judgment of guilty in a criminal case and the punishment that is imposed</DEF> <USAGE>the conviction came as no surprise</USAGE> <BCS>3</BCS> <DOMAIN>law</DOMAIN> <SUMO>Sentencing <TYPE>=</TYPE></SUMO> </SYNSET> </pre>	<pre> </SYNONYM> <ILR><TYPE>hypernym</TYPE> ENG20-01122569-n</ILR> <ILR><TYPE>near_antonym</TYPE> ENG20-01127432-n</ILR> <ILR><TYPE>eng_derivative</TYPE> ENG20-00876567-v</ILR> <ILR><TYPE>category_domain</TYPE> ENG20-06135956-n</ILR> <DEF>(krivično pravo) pravnosnažna, osuđujuća presuda u krivičnom postupku sa dosuđenom kaznom</DEF> <USAGE>Za primanje mita optuženo je 46 lica , a za davanje mita 35. <u>Osudjujuća presuda</u> za primanje mita doneta je u 31 slučaju , dok je za davanje mita osuđeno 30 osoba</USAGE> <BCS>3</BCS> <DOMAIN>law</DOMAIN> <SUMO>Sentencing <TYPE>=</TYPE></SUMO> </SYNSET> </pre>
---	---

Uprkos činjenici da je rad na prilagođavanju pravnih instituta za srpski wordnet bio izuzetno zanimljiv budeći sve vreme stručnu radoznalost, susretala sam se sa problemima koji su proisticali iz činjenice da se američki i naš pravni sistem razlikuju (američki pravni sistem pripada tzv. anglosaksonskom odsono angloameričkom pravu, a naš evropsko-kontinentalnom) te da mnoge pravne termine naš pravni sistem ili ne poznaje ili ih drugačije definiše. U tom smislu sam radeći na pravnim institutima koje ne poznaje naš pravni sistem konsultovala više izvora kako bi prevod bio adekvatan. Jedan takav primer izgleda ovako:

<pre> <SYNSET> <ID>ENG20-06150174-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>assize <SENSE>2</SENSE></LITERAL> </SYNONYM> <ILR>ENG20-06149686-n <TYPE>hypernym</TYPE></ILR> <ILR>ENG20-07928837-n <TYPE>category_domain</TYPE></ILR> <DEF>an ancient writ issued by a court of assize to the sheriff for the recovery of property</DEF> <DOMAIN>law</DOMAIN> <SUMO>Certificate <TYPE>+</TYPE></SUMO> </SYNSET> </pre>	<pre> <SYNSET> <ID>ENG20-06150174-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>uredba <SENSE>2</SENSE></LITERAL> </SYNONYM> <ILR>ENG20-06149686-n <TYPE>hypernym</TYPE></ILR> <ILR>ENG20-07928837-n <TYPE>category_domain</TYPE></ILR> <DEF>sudski nalog (izdavan u prošlosti) administrativnom činovniku za povraćaj imovine </DEF> <NOTE>Not-lexicalized in Serbian</NOTE> <DOMAIN>law</DOMAIN> <SUMO>Certificate <TYPE>+</TYPE></SUMO> </SYNSET> </pre>
--	--

XML element opisa sinseta <NOTE> sadrži belešku koja ukazuje da ovaj koncept u srpskom zapravo nije leksikalizovan. Zanimljivo je uočiti da ni u bugarskom wordnetu koji je takođe razvijan u okviru projekta Balkanet ovaj koncept nije leksikalizovan pa se kao sadržaj elementa <LITERAL> ovog sinseta pojavljuje ‘разпореждане, издавано на шерифа от съда за възстановяване на собственост:1’.

Tokom praktičnog rada najviše sam koristila rečnik (Jovanović, J. i Todorović, S. 2004). S obzirom da opšti rečnici uglavnom ne sadrže stručne pojmove iz oblasti prava, u

radu su korišćeni i mnogi pravni udžbenici (Kovačević Kuštrimović, R. 1997, Jovanović, Lj. 2000, Stanković, G. 1998).

Radi provere ponuđenih rešenja, a u očekivanju da su elektronski tekstovi iz domena prava zastupljeniji od tekstova iz drugih domena, konsultovani su pre svega korpus srpskog savremenog jezika, a ako u njemu potvrda nije pronađena, onda Internet. Pronađene povrde su uključene u elementa <USAGE> XML opisa sinseta, kao što se i vidi iz prvog primera. Pronalaženje potvrda ponuđenih rešenja se pokazalo kao izuzetno osetljiv zadatak, zbog potrebe da se u tekstu pronađe doslovan termin a ne njegova parafraza. Ilustrativan je primer sinseta {potvrđena presuda:1} koja odgovara sinsetu {affirmation:4} sa definicijom “Presuda višeg suda kojom se potvrđuje presuda nižeg suda (kao ispravna)”. Pronađeni primer korišćenja je “Kaznu od 12 godina zatvora Nenadu Brankoviću Vrhovni sud je preinačio na 15 godina robije, dok je Milu Isailoviću <potvrđena presuda> od tri i po godine zatvora”. Možda se ne bi moglo reći da ovaj primer ne predstavlja potvrdu adekvatnosti termina *potvrđena presuda*, ali se mora primetiti da se u primeru ne koristi termin čija je sintaksička struktura *pridev+imenica*, već je u pitanju glagol sa svojim subjektom. Zanimljiv je i primer sinseta {speeding ticket:1} sa definicijom “kazna koja se izdaje zbog vožnje brzinom iznad dozvoljene”. Prvo ponuđeno rešenje bilo je *kazna za prekoračenje brzine*, a pretraživanje Interneta je pokazalo da je moguća i druga formulacija *kazna zbog prekoračenja brzine* (na primer, “Dvojica saobraćajnih policajaca naplatila su ministru unutrašnjih poslova Srbije Boži Preleviću kaznu zbog <prekoračenja brzine> na Ibarskoj magistrali”) pa je i taj termin dodat odgovarajućem sinsetu.

4.6. Domen bibliotekarstva i izdavaštva – Ljiljana Macura

Osnovne studije završila sam na Katedri za bibliotekarstvo i informatiku i zaposlena sam u struci. Nakon višegodišnje prakse u različitim tipovima biblioteka, sada obavljam poslove bibliotekara-informatora u naučnoj čitaonici Odeljenja za pružanje informacionih usluga korisnicima Narodne biblioteke Srbije). Iz domena ‘publishing’ (izdavaštvo) obradila sam 62 sinseta koja su povezana sa ontološkim konceptom katalog ‘Book’ (knjiga) i još 20 sinsetova koji se odnose na kataloge, a koji pripadaju različitim domenima. Tokom niza godina, od početka studija bibliotekarstva i informatike, pa do ovog trenutka, neprestano se širi perspektiva iz koje posmatram ove termine.

Pažnju bibliotekara naročito privlači pojam *catalog (catalogue)*. Catalog (catalogue) na engleskom može biti glagol, imenica, a koristi se i u tvorbi prideva. U srpskom jeziku postoji kao više vrsta reči: *katalog, katalogizacija* (imenice), *katalogizirati* (glagol), *kataloški, katalogiziran* (pridevi). Oblici *katalogizirati* i *katalogizirano* više odgovaraju zapadnoj varijanti srpskohrvatskog jezika, mada se redovno koriste i u srpskom jeziku. Često se koriste i izrazi od reči katalog, u opštem, kao i u užem smislu (u bibliotekarstvu i bibliotečkoj delatnosti): *kataloški obraditi/obrađeno, uneti/uneto u katalog, sačiniti katalog*, i tako dalje.

U Prinstonskom wordnetu nalazimo više literala za određeni pojam, a samo jedan odgovarajući na srpskom, ali i obratno. Pomenuto potvrđuju sledeći primeri.

- Imeničkom sinsetu {cookbook:1, cookery book:1} (knjiga sa receptima i uputstvima za spremanje jela) odgovara u Srpskom wordnetu sinset sa samo jednim literalom {kuvar:1}
- Imeničkom sinsetu {order book:1} (knjiga ili sveska u koju se unose narudžbe kupaca, mušterija; obično u više primeraka) odgovara u Srpskom wordnetu sinset sa više literala {knjiga narudžbi:1, knjiga porudžbi:1, knjiga trebovanja:1}
- Imeničkom sinsetu {paperback book:1, paper-back book:1, paperback:1, softback book:1, softback:1, soft-cover book:1, soft-cover:1} (knjiga sa paprinim koricama)

odgovara u Srpskom wordnetu sinset koji takođe ima više literala {broširano izdanje:1, knjiga u mekom povezu:1}

Primer jednog prevedenog sinseta prikazan je na sledeći način :

<pre>< SYNSET > <ID>ENG20-06015176-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>tome <SENSE>1</SENSE> </LITERAL> </SYNONYM> <ILR> <TYPE>hypernym</TYPE> ENG20-06013091-n </ILR> <DEF>a (usually) large and scholarly book knjiga (obično) velika, i za obrazovanog korisnika </DEF> <DOMAIN>publishing</DOMAIN> <SUMO>Book <TYPE>+</TYPE></SUMO> </SYNSET></pre>	<pre>< SYNSET > <ID>ENG20-06015176-n</ID> <POS>n</POS> <SYNONYM> <LITERAL>tom <SENSE>1</SENSE></LITERAL> <LITERAL>sveska <SENSE>1</SENSE></LITERAL> <LITERAL>svezak <SENSE>1</SENSE></LITERAL> <LITERAL>knjiga <SENSE>1</SENSE></LITERAL> </SYNONYM> <ILR> <TYPE>hypernym</TYPE> ENG20-06013091-n </ILR> <DEF>knjiga (obično velika), i za obrazovanog korisnika </DEF> < SNOTE>Uradila Ljiljana Macura, postdiplomac C. Krstev< /SNOTE> </SYNSET></pre>
---	--

Slede primeri pojmova koje nisu leksikalizovani u srpskom jeziku.

- Za {grimoire:1} (priručnik iz crne magije, za prizivanje duhova i demoni) u srpskom jeziku nema odgovarajućeg prevoda, pa u ;
- Ni za {consuetudinary:1, consuetudinal:1} (priručnik koji opisuje običaje određene grupe, posebno rituale monaških redova) ne postoji doslovan prevod u srpskom jeziku, iako bi termin *običajnik* bio dobar kad bi taj pojam u srpskom uopšte bio pozant;
- Sinset {bestiary:1} (srednjevekovna knjiga (obično ilustrovana) poučnog i zabavnog karaktera sa stvarnim i izmišljenim životinjama) možemo prevesti izrazom *zbirka basni*, pošto ne postoji doslovni prevod na srpski.

Što se korišćene literature tiče, iz praktičnih razloga, ali i iz navike najčešće sam konsultovala rečnike u klasičnom obliku (*ESSE, Leksikon JLZ* , Bensonov *Englesko-srpskohrvatski rečnik, Rečnik srpskohrvatskoga književnoga jezika, Simićev Englesko-srpski enciklopedijski rečnik, Vujaklijin Leksikon stranih reči i izraza, Vukičevićev Pravni rečnik : englesko-srpski*). Od rečnika u elektronskoj formi koristila sam Bibliotekarski terminološki rečnik (BTR ONLINE - www izdanje Bibliotekarskog terminološkog rečnika : englesko-srpskog i srpsko-engleskog), kao i sâm WordNet; služila sam se, naravno i Internetom kao opštim izvorom. Smatram da su, još uvek, klasični izvori pouzdaniji kao referentni izvori od Interneta.

5. Zaključak

Srpski wordnet danas ima 14.593 sinseta od koji 2.240 su uradili učesnici u ovom kooperativnom radu. Smatramo da su naši zajednički naponi pokazali da je kooperativan rad na izgradnji Srpskog wordneta moguć, a naše dosadašnje iskustvo, pak, ukazalo potrebu postojanja osobe koja bi obavljala ulogu urednika, sa zadatkom da usklađuje rad svih saradnika, ali i da kontroliše sve obrađene sinsetove pre nego što ovi definitivno uđu u bazu podataka. Treba imati u vidu da je većina saradnika u ovaj zajednički projekat uložila svoje specifično stručno znanje, bogato iskustvo u pronalaženju i korišćenju raznog referensnog

materijala kao i veliki entuzijazam, ali da mnogi od njih, sasvim prirodno, nemaju neophodna znanja, pre svega iz lingvistike, a posebno računarske lingvistike, te je urednički rad neophodan. Smatramo da će planirano veb okruženje za rad na izradi Srpskog wordneta i njegovo konsultovanje koje će biti razvijeno na osnovama ideja opisanih u (Obradović, I. i Stanković, R. 2008) olakšati rad i budućim saradnicima i uredniku.

Automatska izrada wordneta predstavlja temu koja je već duže vremena aktuelna i privlači mnoge istraživače. U više radova koji su bili predstavljeni na konferenciji *Global Wordnet* održanoj januara 2008. godine u Segedinu govori se o iskustvima u automatskoj izgradnji wordneta za pojedine jezike (slovenački, poljski, i druge). Posebno se u radu (De Mello, G. i Weikum, G. 2008) procenjuje vrednost automatski izgrađenih wordneta i zaključuje da su oni korisni u mnogim informatičkim primenama kao i za izgradnju tradicionalnih leksičkih resursa. Treba imati u vidu da se automatska izgradnja Wordneta najčešće zasniva na korišćenju višejezičkih leksičkih i, posebno, tekstualnih resursa u digitalnom obliku. Iskustvo u radu na srpskom wordnetu je pokazalo da su raspoloživi tekstualni resursi u digitalnom obliku za pojedine domene nedovoljni i za ručni način rada, a da bi verovatno bili neupotrebljivi za automatsko generisanje sinsetova iz tih domena. Štaviše, ni pokrivenost određenih domena tradicionalnim leksikonima i priručnicima u mnogim slučajevima nije adekvatna.

Korišćeni rečnici, udžbenici, priručnici i Internet izvori

Anić, Šime, Klaić, Nikola, Domović, Želimir. *Rječnik stranih riječi i izraza*. Zagreb: Sani-Plus, 2002.

Bjelajac, Branko, Vidović, Dane. *Udar na verske slobode*. Beograd: Alfa i Omega, 2001.

Benson, Morton. *Englesko-srpskohrvatski rečnik*. Prosveta : Ljubljana 1980

Benson, Morton. *Englesko-srpski rečnik* [Elektronski izvor]. (1997)

Branković, Tomislav. *Sekte i politika*. Despotovac: Narodna biblioteka "Resavska Škola", 2000.

Bugarski, Ranko. *Uvod u opštu lingvistiku*. Beograd: Čigoja, 1996

Bugarski, Ranko. *Jezici*. Beograd: Čigoja, 1996

Ćosić, Pavle. *Rečnik sinonima i thesaurus srpskog jezika*. Beograd: Komet, 2007

Ćurčić, Božidar. *Razviće životinja*, Naučna knjiga, Beograd. (1985)

Dumanović, Janko, Marinković, Dragoslav, Denić, Miloje. *Genetički rečnik*, Naučna knjiga, Beograd. (1985)

Dorđević, Radmila. *Gramatika engleskog jezika*. Beograd: Izdanje autora, 2002

Đurđević-Stojković, Biljana. *Verske sekte*. Beograd: Narodna knjiga, 2002.

Englesko-srpski srpsko-engleski rečnik sa gramatikom =English-serbian serbian-english Dictionary&Grammar : ESSE. Beograd: Institut za strane jezike, 2005

Grozdanović-Radovanović, Jelena. *Citologija*, Naučna knjiga, Beograd. (1985)

Grozdanović-Radovanović, Jelena. *Histologija*, Beograd. (1980)

Jarak, Mirjana, Govedarica, Mitar. *Mikrobiologija*, Poljoprivredni fakultet, Novi Sad. (2003)

Jemcev, Vsevolod Tihonovič, Đukić, Dragutin. *Mikrobiologija*, Vojnoizdavački zavod, Užice. (2000)

- Jovanović, Ljubiša. *Krivično pravo*. Niš: Pravni fakultet. (2000).
- Jovanović, Jasmina i Todorović, Svetlana. *Rečnik pravnih termina: srpsko – englesko-francuski = Legal Dictionary: English – Serbian = Termes juridiques : francais – serbe*, Savremena administracija, Beograd. (2004).
- Kičić, Miroljub, Krajičanić, Branka. *Medicinska genetika*, Zavod za udžbenike i nastavna sredstva, Beograd. (1989)
- Klaić, Bratoljub. *Rječnik stranih riječi, izraza i kratica*. Zagreb: Državno izdavačko poduzeće Hrvatske, 1951.
- Klajn, Ivan, Šipka, Milan. *Veliki rečnik stranih reči i izraza*. Novi Sad: Prometej, 2006.
- Klajn, Ivan, Šipka, Milan. *Veliki rečnik stranih reči i izraza*. Novi Sad: Prometej, 2007
- Kovačević, Ljiljana. *Bibliotekarski terminološki rečnik: englesko-srpski, srpsko-engleski*. Beograd: Narodna biblioteka Srbije, 2004.
- Kovačević Kuštrimović, Radmila. *Građansko pravo*. Niš: Pravni fakultet, 1997.
- Krim, Kit. *Enciklopedija živih religija*. Beograd: Nolit, 1992.
- Kristal, Dejvid. *Kembrička enciklopedija jezika*. Beograd: Nolit, 1995
- Kristal, Dejvid. *Enciklopedijski rečnik moderne lingvistike*. Beograd: Nolit, 1998
- Krstić, Ljubiša. *Medicinska virusologija*, Grafopan, Beograd. (2005)
- Krunić, Miloje. *Zoologija invertebrata. Deo 1*, Naučna knjiga, Beograd. (1990)
- Leksikon JLZ*. Jugoslavenski leksikografski zavod, Zagreb 1974
- Leon-Dufour, Xavier. *Rječnik biblijske teologije*. Zagreb: Kršćanska sadašnjost, 1969.
- MacMillan English Dictionary for Advanced Learners: International Student Edition*. Oxford: Bloomsberry Publishing Plc, 2002.
- The new Merriam-Webster dictionary*. Merriam-Webster, Springfield. (1989)
- Marinković, Dragoslav, Tucić, Nikola, Kekić Vladimir. *Genetika*, Naučna knjiga, Beograd. (1989)
- Mihailović, Ljiljana. *Upotreba pasivnih glagolskih oblika u savremenom engleskom jeziku*. Filološki fakultet Beogradskog univerziteta, Monografije, knjiga XII, Beograd 1967
- Moskovljević, Miloš. *Rečnik savremenog srpskohrvatskog književnog jezika*. Beograd: Tehnička knjiga-Nolit, 1966.
- Putanec, Valentin *Francusko-hrvatski ili srpski rječnik*. Zagreb: Školska knjiga, 1974.
- Rečnik srpskohrvatskoga književnoga jezika*. Novi Sad : Zagreb, Matica srpska i Matica hrvatska, 1967-1976
- Ristić, Svetomir. Simić, Živojin. Popović, Vladeta. *Enciklopedijski englesko-srpskohrvatski rečnik*. Beograd: Prosveta, 1956.
- Sapir, Edvard. *Jezik*. Novi Sad: Dnevnik, 1992
- Simić, Dušan. *Englesko-srpski enciklopedijski rečnik*. Kragujevac: Centar za naučna istraživanja SANU i Univerziteta : DSP, 2005
- Stanković, Gordana. *Građansko procesno pravo*. Niš: Pravni fakultet. (1998)
- Tešić, Živojin, Todorović Milan. *Mikrobiologija*, Naučna knjiga, Beograd. (1992)

- Vujaklija, Milan. *Leksikon stranih reči i izraza*. Beograd: Prosveta, 1980
- Vujaklija, Milan. *Leksikon stranih reči i izraza*. Beograd: Prosveta, 1986
- Vujaklija, Milan. *Leksikon stranih reči i izraza*. Beograd: Prosveta, 2005.
- Vukičević, Branko. *Pravni rečnik : englesko-srpski sa obrascima pravnih akata : 40.000 terminoloških jedinica*. Beograd : Grmeč-Privredni pregled 2001.
- Živković, D. (urednik): *Rečnik književnih termina*. Beograd: Nolit, 1992.
- Živković, D. *Teorija književnosti sa teorijom pismenosti – priručnik za nastavnike i učenike*, Beograd: Draganić, 2001.

INTERNET IZVORI

- BTR ONLINE (www izdanje Bibliotekarskog terminološkog rečnika : englesko-srpskog i srpsko-engleskog) (<http://btr.nbs.bg.ac.yu/>)
- Danko Šipka - tematska lista za srpski kao strani jezik (<http://www.public.asu.edu/~dsipka/F2.TXT>)
- Gugl (www.google.com)
- Human Genome Project Information
http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml
- Korpus savremenog srpskog jezika (<http://www.korpus.matf.bg.ac.yu/>)
- Krstarica (www.krstarica.com)
- Metak (<http://www.metak.com>)
- Miriam Webster (www.merriam-webster.com/dictionary)
- PhraseBASE (<http://www.phrasebase.com/croatian/languages/index.php?cat=18>)
- PROZ - Radno mesto za prevodioce (<http://srp.proz.com/job?&print=1>)
- Projekat RASTKO - Internet biblioteka srpske kulture (<http://www.rastko.org.yu>)
- Rečnik Filozofskog fakulteta u Novom Sadu (korišćen i preuzet, tokom 2006. godine)
<http://www.ff.ns.ac.yu/elpub/specst/recnik.htm#recnik>
- ST-L (Lista slanja *Srpska terminologija*)
<http://www.staff.amu.edu.pl/~sipkadan/korpus.html>)
- Wikipedia (<http://sr.wikipedia.org> , <http://en.wikipedia.org>, <http://hr.wikipedia.org>)
- Vokabular (<http://www.vokabular.org>)
- Zaštita bilja <http://www.poljoprivreda.info/?oid=9&id=639>

Literatura:

- Bentivogli, L., P. Forner, B. Magnini and E. Pianta. (2004). Revising WordNet Domains Hierarchy: Semantics, Coverage, and Balancing, in COLING 2004 Workshop on "Multilingual Linguistic Resources", Geneva, Switzerland, August 28, 2004, pp. 101-108.
- Christodoulakis, D. N. (ed.). (2004). Design and Development of a Multilingual Balkan Wordnet (BalkaNet IST-2000-29388) – Final Report.

- Fellbaum, C., ed. (1998). WordNet: An Electronic Lexical Database. Cambridge: Mass: MIT Press.
- Krstev C. (2006). Specifični koncepti Balkana u semantičkoj mreži Wordnet. U Zborniku radova "Susreti kultura", Novi Sad, decembar 2004, eds. Ljiljana Subotić et al, pp. 275-285, Novi Sad: Univerzitet u Novom Sadu, Filozofski fakultet.
- Magnini, B. and G. Cavaglia. (2000). Integrating Subject Field Codes into WordNet. In Gavrilidou M. et al. (Eds.) Proceedings of LREC-2000, Second International Conference on Language Resources and Evaluation, Athens: Greece, 31 May - 2 June, 2000, pp. 1413-1418.
- De Mello, G. and Weikum.G. (2008). On the Utility of the Automatically Generated Wordnets. In Proceedings of the Fourth Global WordNet Conference. Syeged: Hungary, January 22-25, 2008, pages 147-161.
- Miller, George A. (1990). Nouns in Wordnet: A Lexical Inheritance System. Journal of Lexicography 3(4): str. 245-264.
- Niles, I. and Pease, A. (2003). Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In Proceedings of the 2003 International Conference on Information and Knowledge Engineering (IKE '03).Las Vegas: Nevada, June 23-26.
- Obradović, I. , Stanković, R. (2008). Softverski alati za korišćenje jezičkih resursa za srpski jezik. Infoteka (ovaj broj)
- Pease, A., and Niles, I. (2002). IEEE Standard Upper Ontology: A Progress Report. Knowledge Engineering Review, Special Issue on Ontologies and Agents. 17, 65-70.
- Stamou, S., Nenadić, G., Christodoulakis, D. (2004). Exploring BalkaNet Shared Ontology towards Multilingual Conceptual Indexing. In Proceedings of the 4th Language Resources and Evaluation Conference (LREC). Lisbon: Portugal.
- Tufiş, D., Cristea, D., Stamou, S. (2004). Balkanet: Aims, Methods, Results and Perspectives. A general Overview. In Special Issue on BalkaNet Project, Romanian Journal on Information Science and Technology, Tufiş, D. (ed.). Bucureşti: Publishing house of the Romanian academy, pages 9-43.
- Tufiş, D. and Svetla K. (2007). Ontology-Supported Text Classification Based on Cross-Lingual Word Sense Disambiguation. In WILF 2007, LNAI 4578, Masulli, F., Mitra, S. and Pasi, G. (eds.). Berlin Heidelberg: Soringer-Verlag, pages 447-455.
- Vossen, P. (2004) Eurowordnet: A Multilingual Databse of Autonomous and Language-Specific Wordnets Connected via an Inter-Lingual-Index, International Journal of Lexicography, 17(2), pages 161-173.
- Vossen, P. (2004) Introduction to the Special Issue on the BalkaNet Project. In Special Issue on BalkaNet Project, Romanian Journal on Information Science and Technology, Tufiş, D. (ed.). Bucureşti: Publishing house of the Romanian academy.