

Univerza
v Ljubljani
Fakulteta
*za gradbeništvo
in geodezijo*



Jamova cesta 2
1000 Ljubljana, Slovenija
<http://www3.fgg.uni-lj.si/>

DRUGG – Digitalni repozitorij UL FGG
<http://drugg.fgg.uni-lj.si/>

V zbirki je izvorna različica izdajatelja.

Prosimo, da se pri navajanju sklicujete na bibliografske podatke, kot je navedeno:

University
of Ljubljana
Faculty of
*Civil and Geodetic
Engineering*



Jamova cesta 2
SI – 1000 Ljubljana, Slovenia
<http://www3.fgg.uni-lj.si/en/>

DRUGG – The Digital Repository
<http://drugg.fgg.uni-lj.si/>

This is a publisher's version PDF file.

When citing, please refer to the publisher's bibliographic information as follows:

Čeh, M., Smole, D., Podobnikar, T. 2012. Izvedba univerzalne ontologije geografskega prostora v podmnožici predikatnega računa prvega reda = Application of universal ontology of geographic space in a subset of the first-order predicate calculus. Geodetski vestnik 56, 1: 57-69.

http://www.geodetski-vestnik.com/56/1/gv56-1_057-069.pdf

IZVEDBA UNIVERZALNE ONTOLOGIJE GEOGRAFSKEGA PROSTORA V PODMNOŽICI PREDIKATNEGA RAČUNA PRVEGA REDA

APPLICATION OF UNIVERSAL ONTOLOGY OF GEOGRAPHIC SPACE IN A SUBSET
OF THE FIRST-ORDER PREDICATE CALCULUS

Marjan Čeh, Domen Smole, Tomaž Podobnikar

UDK: 004:111:659.2:91

IZVLEČEK

Prostorski podatkovni viri, kot so geodetski referenčni sistem, administrativne prostorske enote, naslovi in topografske teme, so v vseh prostorskih podatkovnih zbirkah podlaga za zajem podatkov in položajno umeščanje tematskih podatkov v prostoru. Temeljne strategije geodetske stroke za približevanje potrebam uporabnikov geodetskih podatkov v smislu povezovanja zbirk podatkov so v oblikovanju povezovalnega semantičnega referenčnega sistema v semantičnem spletu ali tako imenovanem spletu 3.0. V tem prispevku so prikazane možnosti za razvoj orodja za enostavnejše in bolj smiselno iskanje ter integracijo na spletu objavljenih prostorskih podatkov. Za rešitev težave je v raziskavah mogoče zaslediti predloge o nadgradnji sedanjih GIS kot sistemov z znanjem, predstavljenim v obliki ontologij. Gre torej za novo generacijo tehnologije GIS, ki jo nekateri imenujejo tudi pametni GIS. Za zdaj takšen GIS obstaja predvsem na teoretični, ne pa tudi na praktični ravni. V tem delu je predstavljena metoda za modeliranje ontologije geografskega prostora v podmnožici predikatnega računa prvega reda. Izdelano semantično omrežje prostora omogoča analize obstoječih zbirk podatkov za namene integracije v okolju porazdeljenih informacijskih sistemov. Naša izvedba temelji na metodah strojnega učenja in uporabi programskega jezika prolog.

KLJUČNE BESEDE

ontologija, semantični splet, geografski prostor, logika, predikatni račun

Klasifikacija prispevka po COBISS-u: 1.01

ABSTRACT

Spatial data sources, like the geodetic reference system, administrative spatial units, addresses and topographic maps, serve as a base for geo-referencing to the most of dependant thematic spatial databases. The marketing strategy of the surveying profession towards the users of spatial data infrastructure should be in the design of an integrative semantic reference system to be used within the Semantic Web, or so-called Web 3.0. The main motivation for our research was the representation of possibilities to automate tool development for efficient and more sensible approaches to query information within web-published spatial data. In contemporary research there are several solutions offered as upgrades of basic GIS systems with the knowledge presented in the form of ontologies. Therefore, we are faced with the new generation of GIS technology, which has been named "intelligent GIS". In this article, we present method of modelling the semantic reference system as an application of the ontology of geographic space in the subset of first order predicate calculus. Such a semantic network of geographic space represents the foundation for semantic data analyses and data integration in distributed information systems. Our application is based on the methods of machine learning and use of the Prolog programming language.

KEY WORDS

ontology, Semantic Web, geographic space, logic, predicate calculus

1 UVOD

Prostorski podatkovni viri, kot so geodetski referenčni sistem, administrativne prostorske enote, naslovi in topografske teme, se v vseh prostorskih podatkovnih zbirkah uporabljajo kot podlaga za zajem podatkov in položajno umeščanje tematskih podatkov v prostoru, tako imenovano georeferenciranje. Temeljne strategije, ki se uporabljajo v geodetski stroki za približevanje potrebam uporabnikov geodetskih podatkov v smislu povezovanja zbirk podatkov, so oblikovanje povezovalnega semantičnega referenčnega sistema v semantičnem spletu ali tako imenovanem spletu 3.0 (Berners-Lee et al., 2001; Brickley et al., 2004; Berners-Lee, 2010).

Osnovni cilj raziskave je uporabiti hierarhijo konceptov univerzalne ontologije prostora geografskih razsežnosti (UOPGR) (Čeh, 2003) ter jo predstaviti v programskem jeziku, ki podpira zasnove logike in predikatnega računa. Prostor geografskih razsežnosti predstavljajo prostorski objekti geografske razsežnosti; objekti, ki so večji od človeškega telesa in jih ni mogoče v celoti zaznati z enim zaznavnim dejanjem (Egenhofer, Mark, 1995). Glavne značilnosti ontologije UOPGR so: hierarhična zasnova, neodvisnost od podatkovnih objektov, koncepti so izraženi s simboli, univerzalni obseg modela (saj obravnava celotno domeno prostora geografskih razsežnosti), enostavna razširljivost ter možnost izražanja zapletenih konceptov s sestavljanjem osnovnih konceptov. UOPGR lahko teoretično obravnavamo kot primerno orodje za poizvedovanje po prostorskih podatkih, če vemo, da je zaželen lastnost poizvedovalnih jezikov in uporabniških vmesnikov ta, da uporabnikom omogočajo iskanje ter integracijo informacij, ne da bi poznali imena in strukturo zbirk, v katerih so shranjeni podatki (Podobnikar, Čeh, 2012).

UOPGR je povezava med uporabnikom ter aplikacijami in zbirkami prostorskih podatkov, ki vsebuje iskane koncepte. Uporabnik v postopku iskanja ustreznih zbirk prostorskih podatkov formira poizvedbo z razpoložljivimi koncepti UOPGR-ja. Hkrati lahko določa stopnjo kakovosti ujemanja rezultatov iskanja s poizvedbo (Čeh et al., 2004; 2006; Smole et al., 2011). Takšen koncept je na primer lahko »Objekti aktivnosti izobraževanja«.

V raziskavi smo obravnavali naslednje kataloge in interpretacijske ključe zbirk prostorskih podatkov:

- ključ katastrske klasifikacije (UL SRS, št. 16/74; UL SRS, št. 42/86),
- interpretacijski ključ rabe zemljišč (projekt Zajem, MKGP),
- interpretacijski ključ pokrovnosti zemljišč (projekt CORINE, ES),
- objektni katalog temeljni topografski načrt velikih meril TNN VM 500 (MOL),
- objektni katalog topografska baza večje natančnosti (projekt TOPO5, GURS),
- objektni katalog prostorskih oblik (GURS).

V raziskavi predpostavljamo, da smo dokumente (običajno v formatu HTML) predhodno obdelali tako, da vsak koncept UOPGR-ja »ve«, s kakšno frekvenco ga omenja vsak od metapodatkovnih dokumentov. Za izvedbo takšnega postopka na (pol)samodejen način bi bilo smiselno uporabiti metode odkrivanja zakonitosti v besedilih (»text mining«). Koncepte UOPGR-ja predstavljamo s predikatnim računom prvega reda v programskem jeziku prolog. Uporabnik lahko tako na podlagi izdelane hierarhije konceptov UOPGR-ja interaktivno izvaja osnovna poizvedovanja.

2 METODOLOGIJA

Predlagana metodologija zaobjema opis računalniške inženirske obravnave ontologije ter strojnega učenja in uporabo predikatnega računa prvega reda.

2.1 Računalniška inženirska obravnava ontologije

Za izmenjavo metapodatkov in informacij med različnimi geografskimi podatkovnimi zbirkami je treba razviti primerno metodo za formalno opredeljevanje in strojno predstavljanje znanja o prostoru geografskih razsežnosti. Za pridobivanje semantičnega znanja smo razvili iskalno in integracijsko orodje z metodami strojnega učenja za odkrivanje znanja. Zaželeno je, da je samodejno integracijsko orodje čim bolj »razumno«, tako da lahko samodejno odkrije in razreši čim več semantičnih nesoglasij, poleg tega naj orodje ob morebitnih nerazrešenih nesoglasjih ponudi uporabniku čim več informacij za podporo odločanju. Iskalno in integracijsko orodje mora vsebovati določeno znanje, na primer relacijsko znanje, kot je slovar sopomenk. V domeni računalništva je najpomembnejša metodologija za integracijo zbirk podatkov kot primerjava in integracija shem (Spaccapietra et al., 1992). Ta metodologija se je pozneje začela uporabljati na področju geografskih podatkovnih nizov (Devogle et al., 1998). V geoznanostih so se začele tudi privzemati metode teorije komunikologije, na primer metode primerjanja razmerij (Sester et al., 1998), ter ontologije s področij umetne inteligence (Vet in Mars, 1998).

Odkrivanje znanja v zbirkah prostorskih podatkov kot del metode rudarjenja prostorskih podatkov je predmet številnih raziskav (Keim et al., 1994; Knorr et al., 1997; Leung, 2000; Miller, 2008). Rudarjenje prostorskih podatkov je odkrivanje uporabnih vzorcev in razmerij med podatki v zbirkah prostorskih podatkov. Ta pristop omogoča obravnavo dinamičnih prostorskih pojavov. Za rudarjenje prostorskih podatkov je treba ustvariti konceptualno omrežje za sistematično preiskovanje struktur v obsežnih, raznolikih in dinamičnih zbirkah prostorskih podatkov.

Večina algoritmov za rudarjenje prostorskih podatkov temelji na preučevanju razmerij sosedstva, ker lahko nekateri atributi sosednjih objektov vplivajo na obravnavane objekte. Metode za odkrivanje prostorskega znanja iz podatkov so tesno povezane z osnovnimi značilnostmi sistemov za upravljanje zbirk prostorskih podatkov (SDBMS), zato obstaja težnja po integraciji prostorskih algoritmov s sistemi SDBMS (Ester et al., 1997).

Za izdelavo »razumnega« iskalnega in integracijskega orodja obstaja več uporabnih metodologij:

- izdelava integracijske zbirke podatkov semantičnega znanja iz različnih virov (lokalne konceptualne sheme, integracijske izkušnje, uporabniške trditve);
- pridobitev znanja in izkušenj s strani področnih strokovnjakov s tehnikami ekspertnih sistemov in metodami strojnega učenja (Hayne in Ram, 1990);
- določitev meril semantične podobnosti in količinska predstavitev semantike simbolov z uporabo teorije mehkih množic in teorij dokazovanja v metodah umetne inteligence;
- metoda za samodejno večagentno gradnjo modelov znanja v obliki ontologij iz spletnih tabelaričnih struktur (Pivk, 2005);
- izgradnja semantičnega slovarja, metode opisne logike in predikatnega računa.

V raziskavi smo uporabili pristop strojnega učenja z metodologijo izgradnje semantičnega slovarja, opisne logike in predikatnega računa.

2.2 Strojno učenje in uporaba predikatnega računa prvega reda

Učenje je vsaka sprememba sistema, ki mu omogoča, da isto nalogo opravlja bolje. Rezultat učenja je znanje, ki ga sistem uporabi za reševanje novih nalog. Znanje je lahko (Kononenko, 2005):

- množica zapomnjenih podatkov,
- algoritem za reševanje določenih nalog,
- množica napotkov za bolj učinkovito reševanje nalog.

Strojno učenje lahko opredelimo kot opisovanje ali modeliranje podatkov. Pri obravnavi sistemov za strojno učenje ločimo med učnim algoritmom, ki iz množice podatkov in predznanja tvori novo znanje, in med izvajalnim algoritmom, ki samodejno naučeno znanje uporablja za reševanje novih problemov. Vhod v sistem tvorita predznanje in množica podatkov, izhod pa opis (model, hipoteza, teorija), ki te podatke skupaj s predznanjem opisuje in pojasnjuje. Predznanje običajno tvori kar množica možnih modelov. V tem prostoru modelov (predznanju) bo učni algoritem iskal tistega, ki čim bolj ustreza vhodnim podatkom, hkrati pa upošteva merilo optimalnosti, ki ga bo sistem med iskanjem skušal izpolniti. Predznanje lahko vsebuje tudi začetno hipotezo, ki je lahko približna rešitev problema, ter množico heuristik, ki služijo za usmerjanje iskanja v bolj obetavne dele prostora.

Za predznanje, učne primere in prostor hipotez potrebujemo ustrezno predstavitev. V grobem ločimo naslednje predstavitve znanja (Kononenko, 2005):

- izjavni račun,
- predikatni račun prvega reda,
- diskriminantne in regresijske funkcije ter
- verjetnostne porazdelitve.

Predikatni račun prvega reda v primerjavi z izjavnim računom poleg opisovanja objektov omogoča tudi opisovanje relacij med objekti ter uporabo univerzalno in eksistenčno kvantificiranih spremenljivk. Na področju strojnega učenja se običajno uporablja samo podmnožica predikatnega računa prvega reda, ker bi bilo učenje v neomejenem predikatnem računu preveč zapleteno. Navadno se podmnožica predikatnega računa omeji na Hornove stavke, razširjene na negirane literale. Ta podmnožica ustreza stavkom v programskem jeziku prolog. Podpodročju strojnega učenja, ki se ukvarja z učenjem teorij v podmnožici predikatnega računa prvega reda, pravimo induktivno logično programiranje (Džeroski in Lavrač, 2001).

Lastnosti predikatnega računa prvega reda, predvsem možnost opisovanja razmerij med objekti, omogočajo modeliranje konceptov hierarhije UOPGR-ja in njihovih medsebojnih odnosov, kar smo izvedli v nadaljevanju. Omenjeni prostodostopni program deluje na podlagi metod induktivnega logičnega programiranja. Predlagana metoda, podatki, rezultati in ugotovitve delujejo v programskem okolju »Amzi! Prolog + Logic Server«.

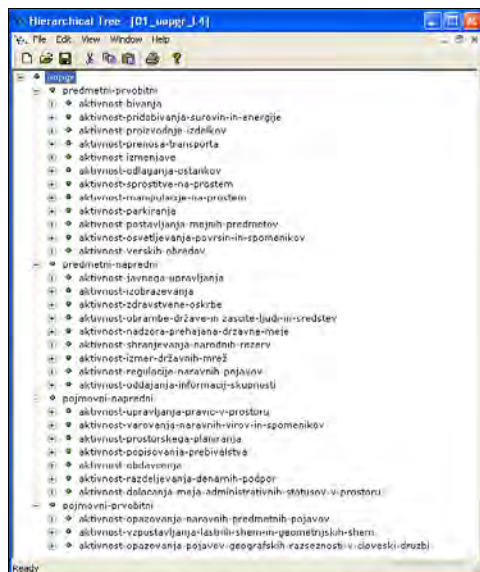
3 REZULTATI

Ročno »pomensko bogatenje« prostorskih zbirk podatkov s koncepti UOPGR-ja je precej zamudna naloga. Razlog za prehod z ročnega na (pol)samodejno indeksiranje prostorskih zbirk podatkov je tudi dejstvo, da je uspešnost iskalnikov precej odvisna od števila indeksiranih dokumentov. Preteklo delo vključuje ročno opisovanje šestih pomembnejših slovenskih zbirk prostorskih podatkov s koncepti UOPGR-ja, kar je podrobneje opisano v Čeh (2003). Celotno hierarhijo konceptov UOPGR-ja sestavljajo koncepti na petih ravneh.

V postopke, ki vključujejo velike količine podatkov, uvajamo računalniško obdelavo. Rezultati te raziskave obsegajo izvedbo predstavitve, analizo semantične ustreznosti UOPGR-ja ter analizo izrazov semantično nezdružljivih konceptov.

3.1 Izvedba predstavitve UOPGR

Najprej je bilo treba razviti orodje, ki omogoča enostavno umeščanje konceptov v pravilna medsebojna razmerja v smislu taksonomije. Razvili smo aplikacijo z imenom »HIT« (slika 1). Uporabnik lahko s tem orodjem vnaša, spreminja ali briše elemente taksonomije. Aplikacija omogoča tudi izpis poljubne hierarhije konceptov v obliki relacije »implies« v formatu prologovih stavkov v tekstovno datoteko. To relacijo, ki označuje podrejenost oziroma nadrejenost med parom konceptov v dani taksonomiji, v prologovih programih običajno označujemo tudi s predikatom »is-child-of« (Bratko, 2001).



Slika 1: Program HIT in UOPGR-drevo

Celotna prologova baza, ki jo izvozimo iz programa HIT, je sestavljena iz dejstev, ki opisujejo razmerja nadrejeni/podrejeni med koncepti s predikatom »implies/2«. Namen takšnega programskega vmesnika je pomagati uporabnikom pri iskanju zbirk prostorskih podatkov, ki

opisujejo/vključujejo iskane in njim bolj ali manj sorodne koncepte. Pojem sorodnost dveh konceptov je v prvi fazi raziskave podana kot razdalja med dvema ali več koncepti. Podajamo krajši izsek iz baze:

- implies(predmetni-prvobitni,uopgr).
- implies(predmetni-napredni,uopgr).
- implies(pojmovni-napredni,uopgr).
- implies(pojmovni-prvobitni,uopgr).
- implies(aktivnost-bivanja,predmetni-prvobitni).
- implies(aktivnost-pridobivanja-surovin-in-energije,predmetni-prvobitni).
- implies(aktivnost-proizvodnje-izdelkov,predmetni-prvobitni).
- implies(aktivnost-prenosa-transporta,predmetni-prvobitni).
- implies(aktivnost-izmenjave,predmetni-prvobitni).
- implies(aktivnost-odlaganja-ostankov,predmetni-prvobitni).
- implies(aktivnost-sprostitve-na-prostem,predmetni-prvobitni).
- implies(aktivnost-manipulacije-na-prostem,predmetni-prvobitni).
- implies(aktivnost-parkiranja,predmetni-prvobitni).
- implies(aktivnost-postavljanja-mejnih-predmetov,predmetni-prvobitni).

V prologu smo napisali program, ki uporabniku omogoča izvajanje poizvedb. Navajamo primer za del poizvedb in odgovorov, pri katerih so najprej navedeni predniki, nato nasledniki:

- ?- predecessor(X,hisa).
- X = vsakodnevno - bivanje ;
- X = uopgr ;
- X = predmetni - prvobitni ;
- X = aktivnost - bivanja ;
- no
- ?- predecessor(hisa,Y).
- no
- ?- predecessor(aktivnost - bivanja,Y).
- Y = vsakodnevno - bivanje ;
- Y = nevsakodnevno - in - sezonsko - bivanje ;
- Y = hisa ;
- Y = blok ;
- Y = stolpnica ;
- Y = stanovanjska - enota ;
- Y = pocitniska - hisa ;

Y = vrtna - hisa ;

Y = pastirski - stan ;

no

Vsebina pregledovalnikov je podana v metapodatkovnem dokumentu oziroma objektnem katalogu zbirke prostorskih podatkov, ki je objavljen v centralnem registru metapodatkovnih opisov (enega takšnih je mogoče najti na spletnih straneh Geodetske uprave RS) ali pa je neposredno na strani, ki vsebuje tudi spletni GIS-pregledovalnik prostorskih podatkov.

3.2 Krajši opis obravnavanih katalogov in interpretacijskih ključev prostorskih podatkov

V nadaljevanju so podani krajši neformalni opisi obravnavanih katalogov in interpretacijskih ključev zbirk prostorskih podatkov.

Opis ključa katastrske klasifikacije

Namen katastrske klasifikacije je določitev zemljišč, ki omogočajo kmetijsko ali gozdno proizvodnjo, ter določitev vrste proizvodne sposobnosti (kulture) in stopnje (razreda) proizvodne sposobnosti posameznih zemljiških parcel oziroma parcelnih delov v okviru katastrskega okraja. Uporabili smo različico klasifikacije v skladu z Zakonom o zemljiškem katastru (1986), po katerem se zemljišče glede na dejansko rabo uvrsti v eno od *skupin* vrst rabe (4 skupine: katastrske kulture, zemljišča pod gradbenimi objekti, zelene površine, nerodovitna zemljišča) in v eno od *vrst rabe* zemljišč (37), ki so določene s Pravilnikom o vodenju vrst rabe zemljišč v zemljiškem katastru (1982).

Opis interpretacijskega ključa kmetijske rabe zemljišč

Ugotavljanje sedanjih kmetijskih rab zemljišč je večnamensko. Osnovni namen je pridobitev podatkov za podporo odločanju za povečanje učinkovitosti gospodarjenja s kmetijskimi zemljišči. Sekundarni namen je nadzor podatkov, navedenih v vlogah kmetijskih pridelovalcev za raznovrstna denarna nadomestila, vezana na zemljišča. Opredelitve posameznih rab zemljišč so namenjene nedvoumnemu prepoznavanju rab iz podatkovnih podlag in razlikovanju med njimi. Raba zemljišč se zajema iz digitalnega ortorektificiranega letalskega posnetka (DOF) v skladu s priloženim šifrantom in opisom opredelitev posameznih kategorij rabe tal. Zajem rabe vključuje zajem vseh zemljišč: kmetijskih, gozdnih, pozidanih in drugih.

Opis interpretacijskega ključa pokrovnosti zemljišč

Osnovni namen projekta Corine je bila priprava digitalne zbirke podatkov o rabi tal in vegetacijskem pokrovu v GIS-okolju, ki naj bi bila podlaga za odločanje in oblikovanje okoljske politike na regionalni in evropski ravni. CORINE land cover je projekt, ki je v Republiki Sloveniji potekal v okviru programa PHARE od septembra 1996 do novembra 1998. Pri delu je bila uporabljena nomenklaturna tabela, razvita za projekte CORINE, ki vsebuje 44 klasifikacijskih razredov (trinivojska struktura) in omogoča primerljivost podatkov s podatkovnimi zbirkami, vzpostavljenimi v drugih evropskih državah.

Opis objektnega kataloga temeljnih topografskih načrtov velikih meril TNN VM 500

Namen izdelave temeljnih topografskih načrtov v večjem merilu (TTN VM) je pridobitev prostorskih podatkov za podporo odločanja na območju mest in naselij ter na drugih območjih, na katerih je predvidena intenzivna izraba prostora. Za našete vrste območij se poleg temeljnih topografskih načrtov v merilu 1 : 5000 izdelajo tudi temeljni topografski načrti v enem od naslednjih meril 1 : 500, 1 : 1000, 1 : 2000, 1 : 2500 po Zakonu o temeljni geodetski izmeri (1974).

Opis objektnega kataloga topografske baze večje natančnosti TBVN

Namen izdelave TBVN in povezave z DOF5 je bila nadomestitev analognega gradiva TTN 5 in 10 z vsebinsko racionalnejšim sistemom, ki ga je lažje vzdrževati in prilagoditi potrebam uporabnikov. Baza TNVB je uporabna kot podlaga za izdelavo kart večjih meril. Zbirka podatkov TBVN se izdela iz več različnih podatkovnih virov in z uporabo več različnih metod. Podatki so razdeljeni na tri ravni (tematsko področje/zbirka, tematska skupina in objekt), model podatkov pa sestoji iz dveh elementov (geometrična in semantična razmerja). Geometrična razmerja predstavljajo pravila za medsebojne odnose med osnovnimi geometričnimi elementi (vozlišče, segment, robni poligon, površina). Podatkovni model vključuje 20 objektov, katerih podatkovni vir predstavlja osem različnih zbirk prostorskih podatkov.

Opis kataloga prostorskih oblik

Namen vzpostavitve kataloga prostorskih oblik je bil podpora pilotnim projektom v okviru projekta ONYX (prostorsko planiranje, upravljanje z nepremičninami in okoljevarstveni vidik prostorskega planiranja). Prostorske oblike so v katalogu prostorskih oblik razvrščene v objektne zbirke, znotraj posamezne objektne zbirke pa še v objektne skupine, ki predstavljajo skupino prostorskih oblik z eno ali več skupnimi lastnostmi. Katalog prostorskih oblik je sestavljen iz kazala in opisa prostorskih oblik. Iz kazala je razvidna celotna vsebina kataloga prostorskih oblik, iz opisa pa podatki posamezne prostorske oblike, ki opredeljujejo njihove lastnosti. Opredelitev prostorskega objekta je izvedena z več elementi, in sicer nazivom objekta, mogočim sinonimom, opisom stvarnega prostorskega pojava, geometrično (topološko) opredelitvijo objekta ter s šifro objekta.

3.3 Analiza semantične ustreznosti ontologije UOPGR

Izvedena je bila analiza povezljivosti posameznih katalogov prostorskih podatkov s semantičnim referenčnim omrežjem UOPGR-ja. Določitev razmerij med simboli katalogov, ontologij aplikacij in koncepti ontologije domene UOPGR-ja je izvedena v ustreznih preglednicah (Čeh, 2003). Razmerja so ocenjena z dvema osnovnima semantičnima razmerjema, predikatoma in z negacijo razmerja:

- razmerje *enakovrednih razredov* (ekvivalent) ontologije domene in razredi ontologije aplikacije;
- razmerje *povezave razredov* (dveh ali več konceptov) ontologije domene, tako da v ontologiji aplikacije tvorijo *en sestavljeni razred* (agregat);
- če v UOPGR-ju ni mogoče najti ustreznega koncepta oziroma ustreznega agregata konceptov, je ocenjeno, da razmerje ne obstaja (nezdružljiv).

Za operativno uporabo smo navedena razmerja izrazili kot formule:

ekvivalent = $se_nanaša_na_enakovredni_razred$ [RazredDomene, RazredAplikacije];

agregat = $se_nanaša_na_sestavljene_razred$ [RazredDomene, RazredAplikacije];

nezdružljiv = $se_nanaša_na_nezdružljivi_razred$ [RazredDomene, RazredAplikacije];

V postopku določitve semantičnih razmerij med koncepti zbirke podatkov in koncepti semantičnega referenčnega sistema uporabimo predhodno v UOPGR-ju prevedeni in obogateni katalog obravnavane zbirke podatkov. V katalog uvedemo metapodatke o semantični skladnosti, tri dodatne attribute (ekvivalent, agregat, nezdružljiv), s katerimi označujemo razmerja med koncepti zbirke podatkov in semantičnim referenčnim sistemom (preglednica 1).

Katalog	Metapodatek semantičnih razmerij					Semantična ustreznost UOPGR (%)	
	ekvival. (e)	agregat (a)	nezdružlj. (n)	skupaj (e+a+n)=s	(e + a)	(e + a)/s	(n)
Raba zemljišč - kataster	31	5	6	42	36	86	14
Kmetijska raba zemljišč (projekt Zajem)	34	15	9	58	49	84	16
Temeljni topografski načrt velikih meril (TTN VM 500)	155	56	3	214	211	99	1
Topografska baza večje natančnosti	10	1	5	16	11	69	31
CORINE Land Cover	27	27	5	59	54	92	8
Prostorske oblike (GURS)	95	60	7	162	155	96	4
Skupaj	352	164	35	551	516	94	6
	64 %	30 %	6 %	100 %			

Preglednica 1: Primeri izračuna semantične ustreznosti UOPGR-ja

Semantična razmerja med koncepti UOPGR in koncepti katalogov smo ocenili za šest zbirke prostorskih podatkov. Semantično skladne koncepte vsake zbirke smo sešteli v stolpcu (e+a). Primerjali smo jih s skupnim številom konceptov v katalogu in jih izrazili v odstotkih v stolpcu (e + a)/s.

Povprečno je mogoče 94 % od skupno 551 konceptov iz obravnavanih katalogov prostorskih podatkov izraziti s katalogom znanja ontologije prostora geografskih razsežnosti. Rezultat analize potrjuje visoko stopnjo semantične ustreznosti izdelane ontologije, še posebej, ker ontologija ni bila ustvarjena kot spoj obravnavanih zbirk podatkov - lokalnih ontologij, ampak kot neodvisna zasnova. Dokazano visoka stopnja semantične ustreznosti izdelane ontologije v razmerju do aplikativnih ontologij potrjuje visoko vrednost neodvisne zasnove UOPGR-ja in s tem njene univerzalnosti za domeno prostora geografskih razsežnosti. Zaradi visoke stopnje semantične ustreznosti je UOPGR primerno orodje za ugotavljanje semantične skladnosti konceptov različnih zbirk prostorskih podatkov. Število semantično nezdružljivih konceptov je navedeno v stolpcu (n) in posebej obravnavano v naslednjem razdelku.

3.4 Analiza izrazov semantično nezdružljivih konceptov

Ker določenih konceptov posameznih katalogov zbirke prostorskih podatkov ni mogoče izraziti s simboli konceptov UOPGR-ja, je bila izvedena podporna analiza izrazov v nezdružljivih konceptih obravnavanih zbirke podatkov (preglednica 2).

Rezultati analize razkrivajo, da se v neuvrščeni konceptih najpogosteje pojavlja prostorski izraz zemljišče (12-krat). Sledijo mu prostorski izrazi stavba (7-krat), površina (4-krat), zgradba (3-krat) in objekt (3-krat). Torej sta najbolj dominantna izraza nezdružljivih konceptov *zemljišče* in *stavba*. Izraza ostalo in drugo ne bosta predmet obravnave, ker se ne nanašata na prostor.

V tem razdelku so bili predstavljeni rezultati analiz semantične ustreznosti UOPGR, ki jih bomo v naslednjem razdelku komentirali.

Katalog konceptov	Izrazi v semantično nezdružljivih konceptih					
	stavba	zgradba	objekt	zemljišče	površina	ostalo, drugo
Raba zemljišč – kataster	2	1	1	2	1	
Kmetijska raba zemljišč (projekt Zajem)				7		5
Temeljni topografski načrt velikih meril (TTN VM 500)	1	1				1
Topografska baza večje natančnosti		1	1	1		1
CORINE Land Cover					3	
Prostorske oblike (GURS)	4		1	2		
Skupaj	7	3	3	12	4	7

Preglednica 2: Analiza izrazov semantično nezdružljivih konceptov

4 RAZPRAVA

Praktična uporaba raziskave je prikazana na podlagi izsledkov analize. Izraz *zemljišče* se tako pomensko izkazuje kot sinonim izraza prostor, ki v UOPGR-ju predstavlja izhodišče. Kot izhodišče ontologije je izraz *prostor* izjemno splošen in zato neprimeren za praktično označevanje konceptov. Očitno je pri oblikovanju kataloga kmetijske rabe zemljišč prevladala uporaba besede zemljišče kot sinonima za izraz prostor. V tem katalogu se je izraz zemljišče pojavil kar sedemkrat. Tudi za izraz *površina* bi bilo mogoče sklepati, da je uporabljen za izražanje splošnega koncepta prostora, četudi prek svojega enakozvočnega, homonimnega pomena za koncept *ploskev*. Tako je mogoče izpeljati enakovredno razmerje med koncepti *prostor* – *ploskev* – *zemljišče* oziroma združeno *prostorska ploskev* – *zemljišče*.

Razpravo semantično nezdružljivih konceptov nadaljujemo z drugim najpogosteje zastopanim izrazom v neuvrstljivih konceptih, to je z izrazom *stavba* (7), ki se mu sopomensko, kot sinonima, pridružujeta s skoraj enakim skupnim deležem v neuvrstljivih konceptih izraz *zgradba* (3) in izraz *objekt* (3). Tudi za to skupino izrazov je mogoče izpeljati enakovredno razmerje med koncepti *stavba* – *zgradba* – *objekt*, ki jih ti izrazi predstavljajo. Izraz *stavba* ima v tem razmerju glede na numerične rezultate analize rahlo dominantno vlogo.

Če analizirane izraze in koncepte strnemo z vidika kognitivnega zaznavanja, je mogoče pritrditi,

da je koncept prostora na najbolj splošni ravni sestavljen iz dveh elementov, ki sta:

- ploskev – zemljišče; element, ki ga zaznavamo kot odprt prostor, in
- stavba – zgradba – objekt; element, ki ga v človeški percepciji zaznavamo kot tridimenzionalne kvadre bolj ali manj pravilnih oblik (v odvisnosti od podrobnosti predstavitve oziroma generalizacije) in zato samodejno sklepamo o njihovem umetnem izvoru.

Oba elementa sta predstavitev koncepta prostora, ki ga je mogoče predstaviti kot unijo:

prostor = [ploskev – zemljišče] \cup [stavba – zgradba – objekt],

ki predstavlja izhodišče UOPGR.

Pri opredeljevanju konceptov prostora za praktično uporabo (npr. v zbirkah podatkov) je treba presegati splošnost navedenih izrazov. Če se njihovi uporabi ni mogoče izogniti, jim je treba obogatiti pomen z dodajanjem izrazov, ki imajo manj splošno semantiko oziroma večjo semantično globino. Takšni izrazi so povezani s koncepti človeških aktivnosti v prostoru, kot je podrobneje opisano v Čeh (2003).

5 SKLEP

Upravljanje zbirk prostorskih podatkov v neodvisno porazdeljenih informacijskih sistemih je stvarnost. Učinkovito samodejno povezovanje tovrstnih zbirk podatkov pričakujemo v bližnji prihodnosti. Strukturna medopravilnost sistemov za upravljanje prostorskih podatkov je predmet tehnološkega razvoja, ki že ponuja rešitve v obliki standardov za GIS-e, semantična medopravilnost znanja o prostoru pa je predmet intenzivnih raziskav. Semantični vidik inteligentnega iskanja in integracije informacij vključuje izdelavo sistema znanja za integracijo podatkov. Gre za ohranjanje pomena informacijskih entitet v različnih kontekstih, kar je mogoče obravnavati kot klasifikacijsko nalogo.

Ontologija se šteje za najpomembnejše sredstvo za reševanje semantične raznolikosti informacijskih sistemov. Gre za pristop reševanja problemov z eksplicitnim podajanjem pomena terminov, ki se uporabljajo v informacijskih sistemih. V bistvu so ontologije uporabne za boljšo komunikacijo, bodisi med ljudmi bodisi med računalniki. Medopravilnost med raznovrstnimi računalniškimi sistemi se doseže s prevajanjem in usklajevanjem pomena med različnimi metodami modeliranja, jezikov in programskih orodij. Pri tem se ontologija uporabi kot metapodatkovni referenčni sistem.

Za povezovanje informacij je treba razviti metode, ki omogočajo razrešitev terminoloških in konceptualnih neskladij. Prvi poskusi razvoja primernih metod so temeljili na posameznih primerih zbirk podatkov in dokumentov. Pri tem je postalo očitno, da je treba razviti standardno ontologijo najvišje ravni, kakršno smo uporabili v naši raziskavi.

Uporabili smo opredelitev univerzalne ontologije prostora geografskih razsežnosti (UOPGR, Čeh, 2003), na podlagi katere smo izvajali pomensko iskanje ter integracijo podatkov zbirk prostorskih podatkov. V navedenem konceptualnem omrežju je pomen besed, s katerimi so označeni prostorski objekti v stvarnosti, pojasnjen z uvrščanjem izrazov o objektih v konceptualno

omrežje prostora geografskih razsežnosti. UOPGR vsebuje tudi semantični slovar. Pri izdelavi modela konceptualnega omrežja smo upoštevali izsledke raziskav ontološkega modeliranja in predikatnega računa 1. reda.

Aplikativni rezultati izvedene raziskave kažejo, da je kar 94 % konceptov, ki jih vsebujejo analizirane zbirke prostorskih podatkov, mogoče opisati s koncepti UOPGR-ja. Pri tem velja, do je bil posamezen koncept dane zbirke prostorskih podatkov opisan kot ekvivalent ali pa kot agregat konceptov UOPGR-ja. Ostalih 6 % konceptov je z vidika UOPGR-ja preveč splošnih (npr. »zemlja«, »parcela«, »objekt«) in jih zato brez postopka semantičnega bogatenja ni mogoče zadovoljivo dobro opisati s koncepti UOPGR-ja. Ugotovili smo, da je razviti model UOPGR zelo zanesljiv in dovolj obsežen za namene semantičnega izražanja prostorskih konceptov.

Literatura in viri:

Berners-Lee, T. (2010). *Long Live the Web*, *Scientific American*, 303(6), 80–85.

Berners-Lee, T., Hendler, J., Lassila, O. (2001). *The Semantic Web*, *Scientific American*, 29–37.

Bratko, I. (2001). *Prolog programming for artificial intelligence*. 3rd ed. Harlow (England): Addison-Wesley.

Brickley, D., Guha, R. (2004). *RDF Vocabulary Description Language 1.0: RDF Schema*. W3C Recommendation; pridobljeno 1. 7. 2011 s spletne strani: <http://www.w3.org/TR/2004/REC-rdf-schema-0040210>.

Čeh, M. (2003). *Semantična integracija zbirk prostorskih podatkov*. Založba ZRC, ZRC SAZU Ljubljana.

Čeh, M., Podobnikar, T., Smole, D. (2004). *Geodata - are they accessible and useful?*. V: Toppen, F.(ur.). *AGILE 7th conference on Geographic Information Science, Heraklion, Greece. Conference proceedings*. Heraklion, 2004: Crete University, 789–794.

Čeh, M., Podobnikar, T., Smole, D. (2006). *Semantic similarity measures within the semantic framework of the universal ontology of geographical space*. V: Riedl, A. (ur.). *Progress in spatial data handling : 12th International Symposium on Spatial Data Handling*. Berlin [etc.]: Springer, 417–434.

Devoegele, T., Parent, T., Spaccapietra, S. (1998). *On spatial database integration*. *International Journal Geographic Information Science*, 12(4), Taylor and Francis, London, Velika Britanija.

Džeroski, S., Lavrač, N. (2001). *An introduction to inductive logic programming*. V: Džeroski, S., Lavrač, N. (ur.). *Relational data mining*. Berlin: Springer-Verlag, 48–73.

Egenhofer, M., Mark, D., (1995). *Naive geography*. *COSIT '95, LNCS 988*, Springer Verlag, Berlin, Nemčija.

Ester, M., Kriegel, H., Sander, J. (1997). *Spatial Data Mining: A Database Approach*. *Advances in Spatial Databases, 5th International Symposium, SSD '97, Berlin, Nemčija*.

Hayne, S., Ram, S. (1990). *Multi-user View Integration System (MUVIS): An Expert System for View Integration*. *6th IEEE Conference on data engineering, Los Angeles, ZDA*.

Keim, D., Kriegel, H., Seidel, T. (1994). *Supporting Data Mining of Large Databases by Visual Feedback Queries*. *Proceedings of IEEE 10th International Conference on Data Engineering, Houston, IEEE Comput. Soc. Press, Los Alamitos, California, ZDA*.

Knorr, E., Ng, R., Shilvok, D. (1997). *Finding Boundary Shape matching Relationships in Spatial Data*, *Advances in Spatial Databases. 5th International Symposium, SSD '97, Berlin, Nemčija*.

Kononenko, I. (2005) *Strojno učenje – 2. popravljena in dopolnjena izdaja*, Fakulteta za računalništvo in informatiko, Ljubljana.

Leung, Y. (1997). *Intelligent spatial decision support systems*. Springer Verlag, Berlin, Nemčija.

Miller, H. J. (2008). *Geographic Data Mining and Knowledge Discovery, The Handbook of Geographic Information Science*, J. P. Wilson in A. S. Fotheringham (ur.), Blackwell Publishing Ltd, Oxford, Združeno kraljestvo.

Pivk, A., (2005). *Avtomatska gradnja ontologij iz spletnih tabel*. Doktorska disertacija, Univerza v Mariboru.

Podobnikar, T. Čeh, M. (2012). *Preface V: Podobnikar, T. (ur.), Čeh, M. (ur.). Universal Ontology of Geographic Space:*

Semantic Enrichment for Spatial Data. IGI Global, 2012. 0-304. Web. 12 Mar. 2012. doi:10.4018/978-1-4666-0327-1

Sester, K., Anders, K., Walter, V. (1998). Linking objects of different spatial data sets by integration and aggregation. Geoinformatica, 2(4). Kluwer Academic Publishers, Dordrecht, Nizozemska.

Smole, D., Čeh, M., Podobnikar, T. (2011). Evaluation of Inductive Logic Programming for Information Extraction from Natural Language Texts to Support Spatial Data Recommendation Services. International Journal of Geographical Information Science.

Spaccapietra, S., Parent, C., Dupont, Y. (1992). Model independent assertions for integration of heterogenous schemas. VLDB Journal, Vol. 1, Springer Verlag, Nemčija.

Več, P., Mars, N. (1998). Bottom-up construction of ontologies. IEEE Transactions on knowledge and data engineering, 10(4), IEEE Computer Society.

Zakon o zemljiškem katastru, Uradni list SRS, št. 16/74, UL SRS, št. 42/86.

Pravilnik o vodenju vrst rabe zemljišč v zemljiškem katastru, Uradni list UL SRS, št. 41/82.

Zakon o temeljni geodetski izmeri, Uradni list SRS, št. 16-142/74.

Prispelo v objavo: 4. oktober 2011

Sprejeto: 16. februar 2012

dr. Marjan Čeh, univ. dipl. inž. geod.

UL FGG - Oddelek za geodezijo, Jamova 2, SI-1000 Ljubljana

e-pošta: marjan.ceh@fgg.uni-lj.si

mag. Domen Smole, univ. dipl. inž. geod.

DFG Consulting, Pivovarniška ulica 8, SI-1000 Ljubljana

e-pošta: domen.smole@dfgcon.si

doc. dr. Tomaž Podobnikar, znanstveni svetnik, univ. dipl. inž. geod.

ZRC SAZU - Inštitut za antropološke študije, Novi trg 2, SI-1000 Ljubljana in

UL FGG - Oddelek za geodezijo, Jamova 2, SI-1000 Ljubljana

e-pošta: tomaz.podobnikar@fgg.uni-lj.si