

# PROŠIRIVANJE UPITA U GEOLOŠKOM INFORMACIONOM SISTEMU SRBIJE (GEOLISS)

Ranka Stanković, Ivan Obradović, Olivera Kitanović  
Univerzitet u Beogradu, Rudarsko-geološki fakultet

**Abstract:** Mineral resources exploitation today increasingly relies on new tools and techniques offered by information technology, among which GIS technology takes a prominent place. Various types of geological datasets can be stored in GIS geodatabases and linked to regional maps and detailed maps of exploration prospects or active mines. Reserve estimates, annual planned production, or cost per ton statistics can be linked to prospect or mine locations. Mining companies also use GIS technology to actively monitor the environmental impacts that may be caused by their activities. The Geological information system of Serbia (GeolISS) has been developed by the Faculty of Mining and Geology in Belgrade University, using ESRI ArcGIS technology. GeolISS is designed to function as a personal geodatabase and SDE enterprise geodatabase on MS SQL server 2000, and many institutions in Serbia are now using this system as it has proved to be a substantial and reliable foundation for digital archiving, query, retrieving, analysis and visualization of geological data. Efficient exploration of such a valuable resource, in view of the diversity and amount of archived data, including textual, is a complex task. Geodatabase queries can be substantially improved by using various lexical resources, such as morphological dictionaries and a geologic dictionary. In this paper we present an improvement of queries within GeolISS using resources and tools developed within the Human Language Technology (HLT) Group at the University of Belgrade. These resources are integrated through a web service WS4QE, and enable semantic and morphological expansion of queries, the latter being very important in highly inflective languages, such as Serbian. The initial results obtained by coupling WS4QE and GeolISS show that resources and tools developed within the HLT Group can substantially improve query results for the GeolISS geodatabase. Further integration of developed lexical resources, namely the spatial multilingual database of named entities, can be used for multilingual symbolization of maps.

**Ključne reči:** GeolISS, proširenje upita, GIS

## 1. GeolISS: Geološki informacioni sistem Republike Srbije

Eksploatacija mineralnih sirovina danas sve više koristi metode, tehnike i alate koje nudi informaciona tehnologija. Među njima, geografski informacioni sistemi (GIS), namenjeni prikupljanju, ahiviranju, upravljanju i prezentiranju podataka koji su prostorno, odnosno geografski referencirani zauzimaju posebno značajno mesto. GIS aplikacije omogućavaju korisnicima da kreiraju upite, analiziraju prostorne informacije, uređuju podatke i karte, i prezentiraju rezultate svih ovih operacija. Pomoću njih se mogu kreirati regionale karte i detaljne karte perspektivnih ili aktivnih površinskih kopova, povezane sa različitim

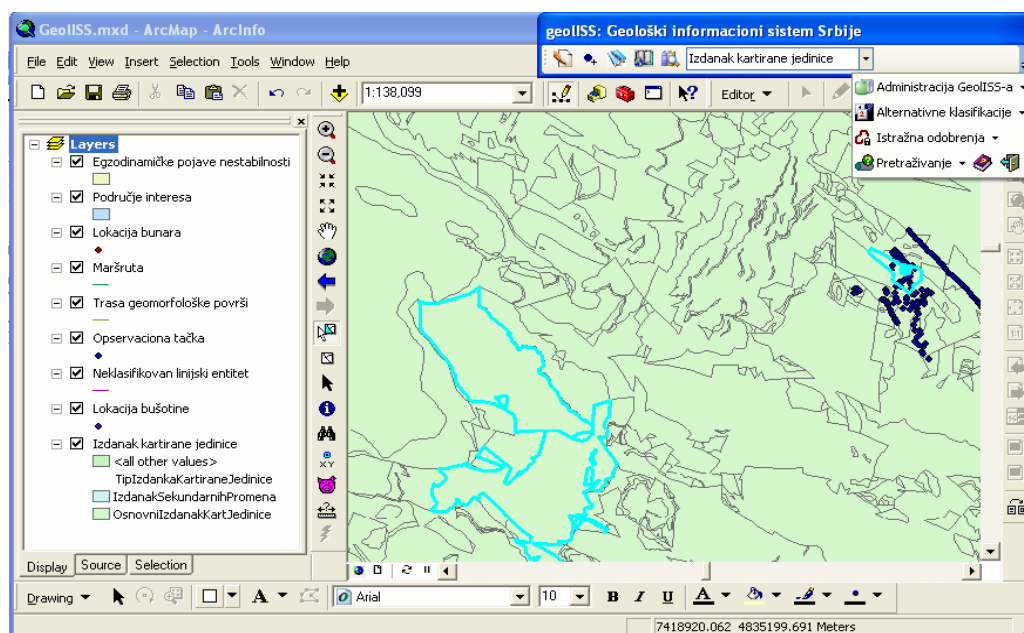
tipovima geoloških podataka značajnim za eksploataciju. Procena rezervi, planirana godišnja proizvodnja ili statistički podaci o troškovima eksploatacije po toni mineralne sirovine mogu se povezati sa lokacijama na karti. Pomoću GIS tehnologije mogu se pratiti pozicije opreme: bagera, traka, vodosabirnika, bunara, zatim stanje radova na etažama, odlagalištima jalovine, deponijama uglja itd. Rudarske kompanije danas aktivno koriste GIS tehnologiju i za aktivno praćenje uticaja svojih aktivnosti na životnu sredinu.

Projektovanje Geološkog informacionog sistema Republike Srbije (GeolISS) finansirano je od strane Ministarstva za zaštitu životne sredine Republike Srbije, a realizovao ga je tim Rudarsko-geološkog fakulteta Univerziteta u Beogradu u periodu od 2005-2006 godine [1]. Realizacijom GeolISS-a stvorena je solidna osnova za digitalno arhiviranje, pretraživanje, analizu i vizuelizaciju geoloških podataka, i njihovo intranet i internet publikovanje [2]. Tokom 2007. godine Ministarstvo za zaštitu životne sredine je finansiralo 28 geoloških projekata u kojima su brojne institucije učestvovala u prikupljanju geoloških podataka za GeolISS. Obradivači su prikupljene podatke unosili u GeolISS replike, koje su potom prenete u centralnu bazu. GeolISS postaje standard za arhiviranje geoloških podataka i kao obavezan alat se koristi za geološke projekte finansirane od strane Ministarstva za zaštitu životne sredine.

GeolISS je implementiran u okviru ESRI, ArcGIS tehnologije kao skup .Net klasa, te predstavlja ekstenziju ArcGIS-a [3]. Sistem podržava distribuirani rad korisnika na udaljenim lokacijama i na terenu, za koji se pripremaju replike u *personal geodatabase* formatu, dok se kao centralni repozitorijum koristi *enterprise ArcSDE geodatabase* u okviru MS SQL Server 2000. GeolISS korisnicima omogućava interaktivno kreiranje upita, analizu prostornih i ostalih podataka, ažuriranje podataka, karata i prikaz rezultata. Modeliranje podataka je inspirisano različitim geološkim modelima [4], među kojima se izdvajaju GeoSciML shema za razmenu geoloških podataka koju koristi CGI IUGS-a ([www.seegrid.csiro.au](http://www.seegrid.csiro.au)), NGMDB implementacija USGS-a i CGS-a ([www.nadm-geo.org](http://www.nadm-geo.org)), kao i modeli predloženi od internacionalne organizacije za standardizaciju geografskih informacija (Geomatics project, ISO TC211, <http://www.isotc211.org>).

Logički okvir implementacije GeolISS-a je zasnovan na pet osnovnih klasa: konceptu, opservacijama, opisu, prostornim entitetima i metapodacima. **Koncept** predstavlja jezgro GeolISS-a preko koga se vrši validacija, klasifikacija i specifikacija vrednosti svojstava opserviranog i interpretiranog, a implementiran je kao agregacija geoloških rečnika, petrološke i mineraloške klasifikacije, geohronološke skale i leksikona kartiranih jedinica. **Opservacije** implementiraju zapis rezultata opserviranih i/ili merenih svojstava, a mogu biti izraženi kvalitativno, kvantitativno, fotografijom, skicom ili geometrijom. **Opis** je implementiran kao instanca opservacija i interpretacija, odnosno, to je kolekcija svojstava koja karakterišu neku geološku pojavu. Tabele opisa osnovnih geoloških entiteta (kartirane jedinice, litološkog sastava i geološke strukture) sadrže zajedničke attribute koji dopuštaju eksplicitno definisanje svrhe i konteksta opisa. Svrha opisa je opis instance ili standardni opis, a kontekst opisa određuje domen prostora na koji se taj opis odnosi. Opis hidrogeoloških pojava i pojava i ležišta mineralnih sirovina je

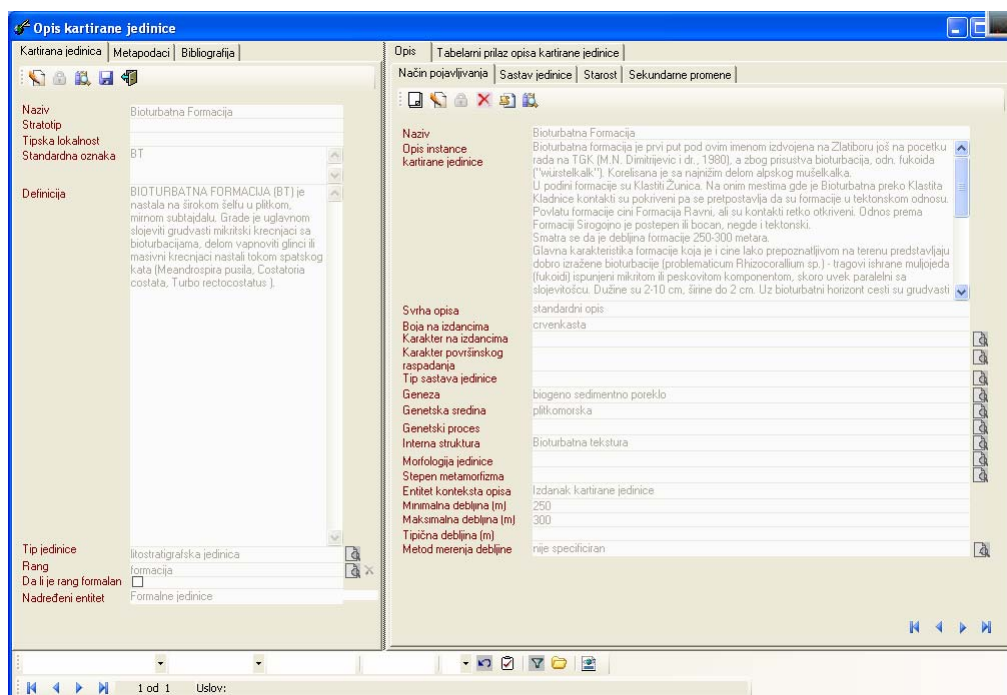
omogućen kroz unos kvalitativnih sadržaja i kroz specifikaciju kvantitativnih, merenih vrednosti. Egzodinamičke pojave nestabilnosti (klizišta, odroni, blatišta i sl.) implementirane su kao katalog sa mogućnošću opisa osnovnih svojstava, uz mogućnost praćenja evolucije pojave u vremenu. **Prostorni entiteti** su lokacije opservacija i opservirani ili interpretirani geološki entiteti u realnom geografskom prostoru. Geometrijski su implementirani kao tačka, duž ili poligon, što omogućava vizuelizaciju geoloških karakteristika u GIS-u i njihovo kartografsko predstavljanje. **Metapodaci** evidentiraju izvore podataka, povezuju ih sa bibliografijom, osobom, institucijom i projektom za koji se unose podaci, datum unosa i tip obrade podataka. **Relacije** su implementirane kao semantički čvorovi koji obezbeđuju povezivanje tematskih ili logički srodnih grupa podataka. Klase relacija obuhvataju direktne relacije, intrarelacije, relacije opservacija i relacije metapodataka.



Slika 1. Selektovanje objekata u GeolISS okruženju

Alati za upravljanje podacima u GeolISS-u su ekstenzije ArcGIS-a, namenski projektovane i implementirane za rad GeolISS bazom [5]. Na slici 1 je prikazano radno okruženje GeolISS-a, integrisano u ArcMap (deo ESRI ArcGIS-a), sa listom izabranih slojeva (lejera) prostornih podataka na levoj strani i odgovarajućom kartom na desnoj. U gornjem delu slike je prikazana paleta alata koja omogućava izbor odgovarajućeg alata za upravljanje podacima. Korisnik može da izabere jedan ili više objekata sa karte i koristeći GeolISS paletu alata otvori odgovarajući interfejs za opis prostornih podataka. Prethodno izabrana geološka jedinica je prikazana na slici 2. Svaki tip geološkog prostornog podatka ima odgovarajuću formu interfejsa za upravljanje podacima, uz mogućnost dodavanja raznih podataka vezanih za njega. Kada su u pitanju podaci o bušotinama u ugljenom

basenu, tipično se unose osnovni podaci o bušotinama, podaci o litologiji, stratigrafska starost, potom različiti tipovi merenja, tehničke, elementarne i analize pepela, geochemijske i geofizičke analize i sl.



Slika 2. Podaci o izabranom prostornom objektu

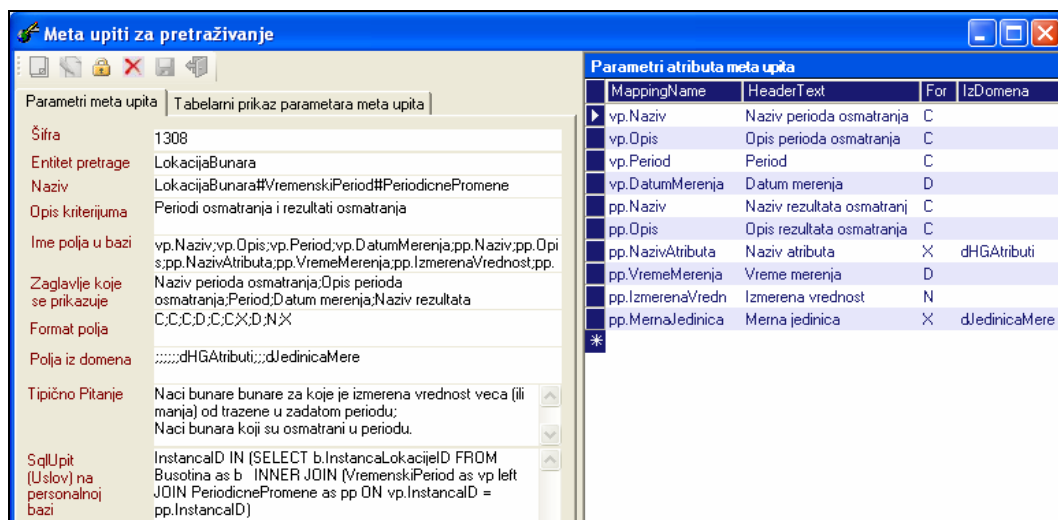
Softversko rešenje interfejsa je implementirano u okruženju MS Visual Studio 2005 i podržava upravljanje podacima i u personalnom i u serverskom izdanju baze (Enterprise geodatabase; SDE i MS SQL Server 2000 i Personal geodatabase). Struktura interfejsa je podređena fleksibilnosti i jednostavnosti korišćenja, pri čemu, osim predviđenog rada u okviru ArcGIS-ArcMap okruženja, GeolISS može da se koristi i kao nezavisna aplikacija, ali bez mogućnosti vizuelizacije prostornih podataka.

## 2. Pretraživanje u GeolISS-u

Imajući u vidu raznovrsnost i količinu podataka pohranjenih u GeolISS-u, jasno je da je efikasno pretraživanje tako važnog izvora geoloških podataka od velike važnosti, tako da je ovom problemu u implementaciji posvećena naročita pažnja. Prostorni upiti se razlikuju od SQL upita jer omogućuju korišćenje geometrijskih tipova podataka (tačka, linija, poligona), i uzimaju u obzir prostorne veze među njima, kao što su udaljenost, razdvojenost, presecanje, dodirivanje, ukrštanje, preklapanje, sadržanje, kao i izračunavanje dužine i površine. Pretraživanje informacija u GeolISS-u je omogućeno na više nivoa.

- Pretraživanje ugrađeno u forme interfejsa
- Pretraživanje prostornih elemenata korišćenjem GeolISS meta upita
- Podrška simbolizaciji spajanjem prostornih i tematskih (atributnih) tabela
- Kreiranje SQL upita korišćenjem GeolISS alata za upravljenje upitima
- Kombinovanje ArcMap funkcija za pretraživanja prostornih objekata sa GeolISS formama.

Pretraživanje prostornih elemenata koje se bazira na unapred pripremljenim obrascima upita (meta upitima) se najčešće koristi. Parametri za upite su unapred definisani u bazi, ali je ostavljena mogućnost da korisnik može da ih dopuni i da po potrebi doda nove definicije za pretraživanje, odnosno nove meta upite. Trenutno postoji 46 meta upita za deset tipova prostornih objekata, pri čemu svaki od ovih meta upita ima više parametara atributa. Na slici 3 je prikazana forma sa jednim meta upitom, na kom se definiše klasa prostornih objekata (u ovom slučaju bunar), naziv i opis kriterijuma za pretraživanje, tipična pitanja koja odgovaraju meta upitu, i lista parametara atributa meta upita. Za svaki atribut meta upita se definiše ime atributa u bazi kojim se vrši mapiranje, potom zaglavlje koje se prikazuju kada se kreira konkretan upit, tip polja sa kodiranim vrednostima za niske karaktere ‘C’, za brojeve ‘N,I’, datum ‘D’, i ako je u pitanju vrednost iz unapred definisanog domena ‘X’, sa nazivom tog domena u susednoj koloni.



Slika 3. Interfejs za upravljanje meta upitima

Na slici 4 je prikazan karakteristična forma za pretraživanje prostornih objekata na kojoj je kao entitet koji se pretražuje izabrana lokacija bunara, ali to može biti i bilo koji drugi prostorni element (bušotina, lokalni stub, kartirana jedinica, područje interesa, trasa geološke površi, egzodinamička pojava, ...). Potom se iz padajuće liste bira kriterijum po kom se pretražuje, pa zatim atribut (polje) po čijem sadržaju se definiše kriterijum. Može se kombinovati više kriterijuma sa Bulovim operatorima (i/ili, tj. AND/OR). U donjem

delu forme je prikazan WHERE deo SQL upita za napredne korisnike kojima je SQL kao upitni jezik blizak, pa oni mogu dalje dopunjavati ili menjati predefinisani upit.

Slika 4. Interfejs za kreiranje i izvršavanje upita za prostorne podatke

Nakon startovanja upita, otvara se odgovarajući sloj (lejer) sa istaknutim prostornim objektima koji zadovoljavaju postavljeni kriterijum. Dalje je moguće automatski otvoriti formu interfejsa sa pronađenim objektima, vršiti dalje filtriranje i koristiti forme na uobičajen način.

### 3. Proširivanje upita u GeolISS-u

Obzirom na veliku količinu tekstualnih podataka koji se nalaze u GeolISS-u, i ovde se pojavljuje tipičan problem postavljanja upita nad tekstualnim podacima. Ovaj problem proističe iz činjenice da korisnik ovakve upite najčešće zadaje navođenjem jedne ili više niski karaktera (reči odnosno oblika reči) povezanih i/ili operatorima, ali ga u suštini interesuju dokumenti ili zapisi vezani za određeni koncept. Koncepti se često mogu označavati sa više različitih reči, odnosno sinonima (npr. klizište, nestabilna padina, ruč), pa jedan deo problema leži u tome što korisnik, prilikom formulacije upita, može propustiti da u upit uključi neke od reči kojima se koncept označava. Ovaj deo problema se može efikasno rešiti semantičkim proširenjem upita, odnosno dodavanjem reči (sinonima) koje je korisnik propustio da navede. Drugi deo problema prilikom izbora reči leži u tome što ista reč može imati različite morfološke oblike (klizište, klizišta, klizištu, itd.) pa pored osnovnog oblika ključnih reči u upite tada često treba uključiti i njih.

Jezički resursi pružaju mogućnosti za sistematično rešavanje izloženih problema semantičkog (sinonimi) i morfološkog (oblici reči) proširenja upita nad tekstualnim podacima. Pored jezičkih resursa, kao što su morfološki rečnici, Grupa za jezičke tehnologije Univerziteta u Beogradu razvila je i metodologiju za integraciju različitih tipova jezičkih resursa [6], a zatim i odgovarajuće alate koji omogućavaju njihovo korišćenje u različite svrhe [7]. U GeolISS-u se tekstualni podaci uglavnom vezuju za opservacije, interpretacije i dnevnik sa terena. Semantičko proširenje upita nad ovim podacima se zasniva na Rečniku geoloških termina (sa sinonimima i engleskim ekvivalentnim terminima), razvijenom u sklopu GeolISS-a, a morfološko na sistemu

morfoloških rečnika Grupe za jezičke tehnologije. Da bi se korisniku obezbedila što veća fleksibilnost u izboru niski od kojih će se formirati konačni upit, GeolISS je integrisan sa jednim od softverskih proizvoda Grupe za jezičke tehnologije, veb servisom za ekspanziju upita sistema WS4QE (Web Service for Query Expansion), koji daje korisniku mogućnost da upit proširi semantički, morfološki, ali i na još jedan jezik (engleski).

Da bismo ilustrovali značaj problema pretraživanja tekstualnih podataka u GeolISS-u kao primer ćemo navesti rezultate pretraživanja za geološke jedinice prema ključnoj reči *glina* u polju sa opisom. Upit je urađen dva puta: jednom bez proširenja i drugi put sa morfološkim proširenjem uz pomoć WS4QE. Za reč *glina* u proširenom upitu WS4QE je generisao morfološke oblike '*glina, glinama, glinom, glinu, glini, gline*', pa su svi ovi oblici uključeni u WHERE deo SELECT upita. Karakteristični delovi neproširenog i proširenog upita su:

- KlasifikacioniTermin IN (SELECT InstancaID FROM KartiranaJedinica WHERE Opis LIKE '%glina%')
- KlasifikacioniTermin IN (SELECT InstancaID FROM KartiranaJedinica WHERE Opis LIKE '%glina%' OR Opis LIKE '%glinama%' OR Opis LIKE '%glinom%' OR Opis LIKE '%glinu%' OR Opis LIKE '%glini%' OR Opis LIKE '%gline%')

Slika 6 prikazuje na levoj strani rezultate za originalni, neprošireni upit, sa izdvojenom i označenom 51 geološkom jedinicom, a na desnoj strani rezultate proširenog upita, sa 118 geoloških jedinica, dakle dvostruko više.

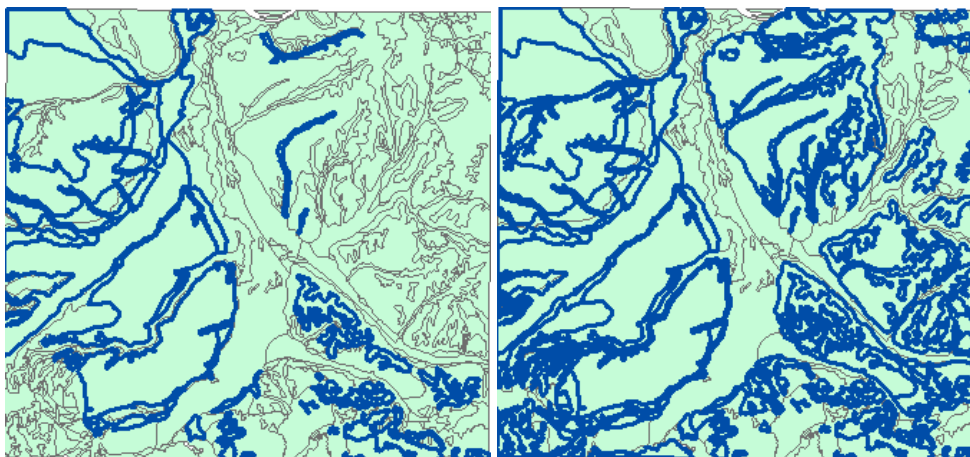


Figure 6. Rezultat originalnog i proširenog upita

Na sličan način moguće je vršiti i semantička proširenja upita uključivanjem sinonima, koristeći informacije iz geološkog rečnika, kao za već navedeni primer za klizište: nestabilna padina, ruč.



#### 4. Zaključak

U radu je pokazano kako se leksički resursi za srpski jezik, razvijeni u okviru Grupe za jezičke tehnologije, a integrisani kroz veb servis WS4QE, mogu efikasno koristiti za proširenje upita u Geološkom informacionom sistemu Republike Srbije. Početne analize efekata proširivanja upita pokazale su da se na ovaj način mogu dobiti znatno bolji rezultati, što je i ilustrovano jednim primerom u kome je uz morfološko proširenje dobijen dvostruko veći broj geoloških jedinica nego kada je upit postavljen bez ovog proširenja. Sličan efekat može se postići semantičkim proširenjem upita kada su u pitanju geološki pojmovi koji se označavaju pomoću više sinonima. Dalja istraživanja mogućnosti proširivanja upita u GeolISS-u uz pomoć jezičkih resursa motivisana su, između ostalog, i činjenicom da GeolISS postaje standard za arhiviranje geoloških podataka u Srbiji i da se kao obavezan alat koristi za geološke projekte finansirane od strane Ministarstva za zaštitu životne sredine. Ova istraživanja biće usmerena pre svega na dalji razvoj resursa koji su već integrisani, ali i integrisanje drugih resursa kao što su već postojeća višejezična baza imenovanih entiteta (koji obuhvataju i toponime) čime će se omogućiti višejezično obeležavanje karata.

#### Bibliografija

- [1] Blagojević B., Trivić B., Banjac N., Stanković R., Nikolić V. (2005). "Razvoj Geološkog informacionog sistema Republike Srbije", *Zbornik radova 14. kongresa geologa Srbije i Crne Gore*, Novi Sad 18-20.oktobar 2005, s.219-224.
- [2] Blagojević B., Trivić B., Stanković R., Banjac N. (2008). "Short note about implementation of Geologic information system of Serbia", *Zapisi Srpskog geološkog društva*, Srpsko geološko društvo, Beograd.
- [3] ESRI Developer network (<http://edn.esri.com>)
- [4] Richard, S.M. (2003). Geologic map database implementation in the ESRI Geodatabase environment, in Soller, D.R., ed., *Digital Mapping Techniques 03—Workshop Proceedings*, U.S. Geological Survey Open-File Report 03-471, p.169-183, accessed at <http://pubs.usgs.gov/of/2003/of03-471/richard2/>
- [5] Stanković R. (2008). "Improvement of geodatabase queries within GeolISS", *Review of National Centre for Digitization*, Matematički fakultet, Beograd. (u štampi)
- [6] Krstev, C., Vitas D., Stanković R., Obradović I., Pavlović-Lažetić G. (2004). "Combining Heterogeneous Lexical Resources", *Proceedings of the Fourth International Conference LREC*, Lisbon, Portugal, May 2004, vol. 4, pp. 1103-1106.
- [7] Krstev, C., Stanković, R., Vitas, D., Obradović, I. (2006). "WS4LR: A Workstation for Lexical Resources", *Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC 2006*, Genoa, Italy, May 2006, pp. 1692-1697.