

## ŽIVOTNI CIKLUS POVEZANIH PODATAKA KROZ RAZVOJ SKUPA PODATAKA VOZILA U POKRETU LINKED DATA LIFECYCLE OF FLOATING CAR DATASET

Mirko Spasić

**REZIME:** Posljednjih godina, principi povezanih podataka znatno doprinose ostvarivanju vizije Semantičkog veba. Javno dostupni podaci, usklađeni sa ovim principima, daju ogroman potencijal razvoju aplikacija različitih namena, jer kao takvi, uklanjaju barijere nametnute ograničenim pristupom podacima, kao i različitim i nestandardnim formatima. Jednom objavljeni, povezani podaci otpočinju svoj životni ciklus, koji obuhvata njihovu konstantnu transformaciju, u smislu povezivanja i spajanja sa drugim podacima, klasifikacije i obogaćivanja, analize kvaliteta, popravke i evolucije, a uslovljen korišćenjem u različitim aplikacijama. U ovom radu, prikazan je životni ciklus geoprostornog skupa podataka vozila u pokretu, kroz sve faze unapređivanja, uključujući transformacije koje ga prave kompatibilnim sa GeoSPARQL standardom, dodaju mu treću dimenziju, i prilagođavaju različitim scenarijima korišćenja. Ovako unapređen skup podataka pruža ogromne mogućnosti za jednostavnu implementaciju složenih GIS aplikacija koje rešavaju suštinski važne probleme iz oblasti analize i upravljanja saobraćajem, što je i demonstrirano razvijenom veb aplikacijom.

**KLJUČNE REČI:** povezani podaci, geoprostorni podaci, saobraćajni podaci, GeoSPARQL standard, životni ciklus povezanih podataka, RDF, SPARQL, GIS

**ABSTRACT:** In recent years, Linked Data principles has contributed significantly towards achieving the vision of the Semantic Web. Linked Open Data represents a huge potential for developing applications from variety of domains by cutting out the barriers imposed by restricted data access and diverse, non-standard data formats. When published, its Lifecycle starts to spin around, in the constant data transformation, in terms of interlinking, fusion with other datasets, classification and enrichment, quality analysis, repairing and evolution, driven by data exploitation through searching, browsing or exploration. In this paper, Floating Car Dataset Lifecycle is presented, over all phases of transformation that makes it compliant with GeoSPARQL standard, lifts it to 3D, and adapts it to different use cases. The enhanced dataset provides major opportunity for simplifying the implementation of complex GIS applications that solve important problems from the area of traffic analysis and traffic management, as demonstrated by developed web application.

**KEY WORDS:** Linked Data, geospatial data, traffic data, GeoSPARQL standard, Linked Data Lifecycle, RDF, SPARQL, GIS

### 1. UVOD

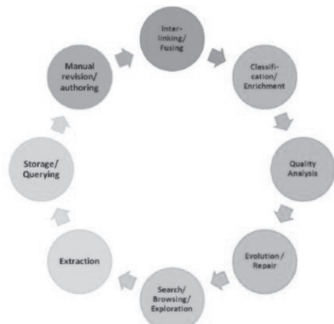
Jedna od glavnih barijera koja onemogućuje korišćenje podataka u njihovoj punoj snazi i kapacitetu je njihova distribuiranost po hiljadama različitih silosa podataka (engl. *data silos*) u kojima su izolovani i nedostupni ostalim korisnicima koji bi od njih imali koristi [1]. Takođe, različiti formati i nepoštovanje standarda mogu dodatno otežati zadatke koji uključuju integraciju informacija i podataka iz različitih izvora [2]. Upravo je olakšavanje ovakvih situacija cilj Semantičkog veba, koji to pokušava da reši obezbeđujući arhitekturu baziranu na okviru (engl. *framework*) za opisivanje veb resursa (engl. *RDF*) i ontologijama, kombinujući je sa tradicionalnim veb tehnologijama (kao što su HTTP i REST) [3]. Njegova glavna ideja je da podaci budu dostupni na vebu, definisani i povezani na način koji omogućava da budu čitljivi i računarima (engl. *machine-readable*), ali ne samo za njihovo prikazivanje, već i za automatizaciju, integraciju i ponovno korišćenje u različitim aplikacijama [3; 4]. Jedan od pragmatičnih pristupa za ostvarivanje vizije Semantičkog veba sadržan je u principima povezanih podataka (engl. *Linked Data*), koji se odnose baš na način njihovog objavljivanja i povezivanja na vebu [5; 6]. U RDF modelu, konkretni resursi predstavljeni su veb identifikatorima (URI), i opisani su navođenjem svojstava i njihovim vrednostima. Model podataka sličan je klasičnom modelu objekti-veze (engl. *entity-relationship model*), i baziran je na ideji da se sve činjenice o resursima mogu iskazati u obliku subjekat-predikat-objekat, tj. tripletima. Subjekt predstavlja resurs koji se opisuje, a predikat neko njegovo svojstvo ili aspekt i povezuje ga sa nekim drugim resursom ili konkretnom vrednošću u objektu. RDF model se najčešće predstavlja grafom, gde su resursi (subjekti i objekti) predstavljeni čvorovima, a predikat označenom ivicom između njih. Serijalizacija ovakvih podataka može biti u različitim formatima: XML,

Turtle, N-Triples, N-Quads, JSON, itd [6]. Baze podataka koje pohranjuju ovakve podatke nazivaju se triplet-skladišta (engl. *triplestores*), i dok su neke izgrađene samo za tu svrhu (npr. Apache Jena [7], RDFox [8], OntoQuad [9]), neke predstavljaju nadogradnju sistema za upravljanje relacionim bazama podataka (npr. OpenLink Virtuoso [10]), ili grafovskim bazama podataka (npr. GraphDB [11], Blazegraph, Neo4J [12]). Sva ova triplet-skladišta, iako mogu imati sopstveni API za pristup podacima, podržavaju upitni jezik SPARQL [13], koji predstavlja standard za manipulaciju podacima u RDF formatu donesen od strane međunarodne grupe za standardizaciju W3C<sup>1</sup>.

U ovom radu, korišćenjem javno dostupnih resursa, opisana je priprema skupa podataka vozila u pokretu (engl. *floating car dataset – FCD*) za željeni region u RDF modelu, a sa ciljem omogućavanja eksploatacije ovih podataka u rešavanju važnih problema kao što su analiza i optimizacija saobraćaja u pametnim transportnim sistemima. Podaci koji bi se koristili u ovim scenarijima postoje, ali su ili nedostupni ili se njihovo korišćenje naplaćuje, tako da rezultujući skup podataka čija se priprema opisuje predstavlja veliki korak u ovom smeru. Takođe, cilj ovog rada je i da postupno prikaže pripremu podataka obavljenu u skladu sa odgovarajućim standardima i prateći generalno prihvaćene metode i tehnike (engl. *best practices*) za svaku od faza transformacije podataka u njihovom životnom ciklusu (engl. *Linked Data Lifecycle*) koje su prikazane na slici 1, i tako doprinese boljem razumevanju svake od njih. Krenuvši od velikog skupa podataka koji sadrži simuliran saobraćaj u nemačkom gradu Lajpcigu u trajanju od dve godine, u procesu njegovog unapređenja i kasnijeg korišćenja kroz razvijenu veb aplikaciju, zastupljene su sve faze koje Auer i ostali razmatraju [6]. Ekstrakcija podataka (engl. *Extraction*) mapira podatke iz nekog drugog formata u RDF, i praćena je njihovim skladištenjem i upitima nad

<sup>1</sup> <https://www.w3.org/>

njima (engl. *Storage & Querying*). Ručne revizije (engl. *Manual Revision & Authoring*) unose pojedinačne promene u podatke, dok povezivanje i objedinjavanje sa drugim izvorima (engl. *Interlinking & Fusion*) ima za cilj dobijanje novih informacija koje su u njima prisutne. Fazom obogaćivanja i klasifikacije (engl. *Enrichment & Classification*) dodaje se struktura višeg nivoa podacima u cilju efikasnije agregacije i upita nad njima. U fazi analize kvaliteta podataka (engl. *Quality Analysis*) mogu se detektovati neki problemi, koji se potom u fazi evolucije i popravke (engl. *Evolution & Repair*) otklanjaju. Faza pretrage, pregleda i istraživanja (engl. *Searching & Browsing & Exploration*) predstavlja eksploataciju tako dobijenih podataka.



Slika 1: Životni ciklus povezanih podataka (slika preuzeta iz [6])

Potpoglavlje 1.1 sadrži opis najznačajnijih sistema koji barataju sa geoprostornim podacima, sa akcentom na one iz oblasti Semantičkog veba. U poglavlju 2, opisan je inicijalni skup podataka, koji je korišćen kao osnova za sve transformacije, dok se u poglavlju 3 prikazuje njegovo unapređivanje. Potpoglavlje 3.1 opisuje transformaciju skupa podataka koja ga usklađuje sa standardom (faza obogaćivanja i klasifikacije). U potpoglavlju 3.2 podacima se dodaje nova dimenzija, kojom podaci postaju trodimenzionalni (faza ekstrakcije). Uklanjanje grešaka merenja u podacima, i njihovo prilagođavanje mapi prikazano je u poglavlju 3.3 (faza analize kvaliteta i faza evolucije i popravke). U potpoglavlju 3.4 izdvojena je mreža segmenata puteva iz putanji vozila (faza obogaćivanja i klasifikacije), i za svaki od njih pronađen odgovarajući iz većeg skupa podataka sa dodatnim informacijama, od kojih su neke materijalizovane dodavanjem u rezultujući skup podataka (faza povezivanja i objedinjavanja). Poglavlje 4 sadrži korisne scenarije i primere korišćenja, implementirane u vidu veb aplikacije koja demonstrira visoku upotrebljivost ovako dobijenih podataka (faza pretrage, pregleda i istraživanja), dok su u poglavlju 5 navedeni glavni zaključci rada.

### 1.1. Opis sistema za rad sa geoprostornim podacima

Mnogi informacioni sistemi sadrže neku vrstu geoprostornih podataka, čija se upotrebljivost ne ogleda samo u prisutnosti geografskih koordinata, tj. vrednosti geografske širine i dužine tačaka i drugih objekata, već i u mogućnosti geoprostornog rezonovanja u smislu izračunavanja odnosa i veza između podataka koje nisu eksplicitno date u samim podacima, već se mogu zaključiti na osnovu njihovih lokacija. Informacioni sistem koji ima za cilj snimanje, čuvanje, manipulaciju, analizu, upravljanje i predstavljanje takvih podataka naziva se geografski informacioni sistem, skraćeno GIS (engl. *Geographical Information System*). Kako se u ovoj oblasti najčešće radi o ogromnim količinama podataka, upravljanje njima razvijano je decenijama unazad kao specifična podoblast upravljanja podacima (engl. *data*

*management*). Mnoge relacione baze podataka (PostgreSQL, IBM DB2, Microsoft SQL Server, Oracle Spatial, MySQL) podržavaju rad sa geografskim podacima [14], čuvajući ih u višedimenzionalnim indeksnim strukturama kao što su prosta B-stabla, ili naprednija R-stabla razvijana baš za ovu potrebu.

Na tržištu postoji mnogo digitalnih mapa i veb servisa koji nude svoje podatke u različitim formatima i u različite svrhe (npr izrada karata, navigacioni softver, usko specijalizirane aplikacije). Neka od njih su Google Maps, Bing Maps, Here WeGo, Apple Maps i dr. Ono što je zajedničko za njih jeste da predstavljaju vlasnički, neslobodan softver, i nameću mnogo ograničenja kako na samo korišćenje servisa, tako i na taj način dobijenih podataka. Jedna od kompanija iz ovog domena je i TomTom [15], koja kao svoj produkt nudi kako mape puteva, tako i navigacioni softver ugrađen u GPS uređaje i mobilne telefone. Broj korisnika njihovih servisa veći je od milion, pa svakodnevno prikupljajući ogromne količine podataka o trenutnim lokacijama, TomTom je u mogućnosti da u svakom trenutku ima informaciju o trenutnom stanju na putevima, preopterećenosti saobraćaja, zastoju zbog radova ili saobraćajne nesreće. Tako prikupljeni podaci koriste se u pametnim transportnim sistemima [16], na primer za rutiranje vozila opremljenih adekvatnim uređajima znatno im štedeći vreme, ili za upravljanje i analizu saobraćaja koja se može iskoristiti u različitim domenima.

OpenStreetMap (OSM) [17] je zajednički projekat velikog broja volontera širom sveta udruženih oko zajedničkog cilja da naprave javno dostupnu digitalnu mapu sveta (OSM skup podataka) koju svaki pojedinac može da proširi saznanjima iz svog okruženja. Kao takva, ona predstavlja pandan Vikipediji u oblasti geografskih podataka kojom je i inspirisana, ali za razliku od nje, korisnici ne uređuju članke, već geografske entitete. Nedostupnost digitalnih mapa i ograničenja nametnuta na njihovu upotrebu motivisala su veliki broj pojedinaca da se uključi i da pomogne rastu ovog skupa podataka. Trenutni broj registrovanih korisnika koji su imali doprinos sadržajima OSM podataka je preko 6.3 miliona. O njegovom ubrzanom razvoju svedoči i preko 3 miliona izmena mapi dnevno, i sam skup podataka koji sadrži preko 7.7 milijardi GPS tačaka i u nekompresovanom OSM XML formatu je veličine preko 1.2 TB. Tako sveobuhvatan, OSM skup podataka vrlo je koristan u različitim scenarijima upotrebe, od pripreme papirnih i elektronskih mapa, pretrage ulica i adresa, do navigacionih aplikacija, pa ga koriste i Facebook, OsmAnd, Foursquare, TomTom, i drugi. Podaci se čuvaju u relacionoj bazi podataka (PostgreSQL), i može im se pristupiti radi postavljanja upita ili izmena. Takođe, podaci se mogu dobiti i u formi izvoza kompletne baze (engl. *database dump*), kao i kasnijih izmena (engl. *changeset*) kojim se raniji izvozi mogu ažurirati i sinhronizovati sa najnovijim stanjem OSM skupa podataka.

Sa ubrzanim razvojem semantičkih tehnologija, raste i potreba za geografskim podacima u njima, pa sve više triplet-skladišta podržava manipulaciju njima [18] (npr Neo4J, OpenLink Virtuoso, Apache Jena, Strabon). Takva podrška ogleda se u mnoštvu funkcija koje barataju sa geoprostornim podacima, na primer računaju rastojanja između dva objekta, ili testiraju da li objekti zadovoljavaju neki geoprostorni predikat (ukršta se, dodiruje, preklapa, itd.). Na samom početku njihovog razvoja, svaki sistem je dodavao sopstvenu geoprostornu podršku, iniciran potrebama i željama svojih korisnika. Sa sve većim brojem različitih sistema sa ovakvom podrškom, aplikacija koje eksploatišu ove podatke i njihovih korisnika, javila se potreba za

sistematizacijom i standardizacijom proširenja SPARQL jezika u ovom aspektu. Međunarodna organizacija za standardizaciju Open Geospatial Consortium (OGC), koja iza sebe ima više od 30 standarda u oblasti geoprostornih podataka, reguliše i ovaj aspekt jezika u vidu GeoSPARQL standarda [19; 20]. On se manifestuje kroz nove tipove podataka, standardizovanu reprezentaciju geoprostornih podataka u RDF okviru kroz nov usaglašen vokabular i osam međusobno nezavisnih topologijskih relacija kojim se mogu opisati svi međusobni odnosi između dve instance ovih tipova. Postoje mnogo benčmarka (engl. *benchmark*) koji testiraju kako performanse ovakvih sistema, tako i njihovu usklađenost sa standardom [18; 21-23].

LinkedGeoData (LGD) [24] je inicijativa da se doda prostorna dimenzija Semantičkom webu, i njen glavni rezultat je RDF baza znanja (engl. *knowlegde base*) nastala od OSM skupa podataka držeći se principa povezanih podataka. Njena veličina direktno odgovara veličini OSM skupa podataka i u broju tripleta iznosi preko 5.1 milijardi. Osim konverzije OSM formata u RDF, ovaj skup podataka sadrži i veze entiteta iz ovog skupa sa drugim geoprostornim RDF skupovima podataka (npr. GeoNames), kao i sa DBpedia skupom podataka, kao najvećim referentnim skupom koji sadrži enciklopedijske podatke iz različitih domena. LGD skup podataka je nastao sa idejom da olakša realne potrebe za integracijom i agregacijom podataka koji u sebi imaju geoprostorni aspekt. Dostupan je za preuzimanje u formi kompresovanih RDF datoteka, ali se upiti nad njim mogu posavljati i preko REST servisa pristupajući SPARQL interfejsima (engl. *SPARQL endpoint*). Kako se OSM podaci konstantno menjaju i dopunjuju, LGD ima modul za sinhronizaciju kojim konvertuje sve kasnije objavljene OSM izmene i održava svoje podatke ažurnim. O korisnosti ovih podataka svedoče mnoge aplikacije različitih domena koje ga koriste (LGD Browser, Map4RDF, Layar, Vicibit, GEM i druge).

## 2. SKUP PODATAKA VOZILA U POKRETU

Kao inicijalna osnova za skup podataka korišćeni su sintetički podaci u RDF formatu dobijeni TomTom-ovim simulatorom [25]. Oni se sastoje od veštačkih putanja vozila na javnim putevima generisanim specijalizovanim algoritmom koji simulira realan saobraćaj na zadatoj teritoriji u određenom vremenskom intervalu. Algoritmu je potrebno obezbediti statističke podatke bazirane na realnom saobraćaju na željenom području, kao i odgovarajuću mapu puteva. Algoritam oponaša ogroman broj GPS uređaja koji prikupljaju podatke o trenutnoj lokaciji i brzini vozila u bliskim vremenskim trenucima (npr. na svakih 5 sekundi) i generiše FCD skup podataka. Glavni entitet od kojih se sastoji je jednokratna putanja nekog vozila po mapi. Svaka od njih sadrži mnogobrojne tačke, tj. drugi entitet određen koordinatama vozila, vrednostima geografske širine i dužine, u tačno određenom vremenskom trenutku, kao i entitetom brzina, predstavljenim vrednostima i jedinicom u kojoj je vrednost data. Veličina generisanog skupa podataka može se zadati unapred kao broj putanja koje treba da budu generisane. Za potrebe ovog rada, simuliran je saobraćaj u i oko grada Lajpciga (Nemačka) u trajanju od dve godine, koji se sastoji od milion putanja sa ukupno više od 126 miliona tačaka, gde rezultujući skup podataka broji više od 889 miliona tripleta.

## 3. UNAPREĐENJA PODATAKA

U ovom poglavlju opisana su sva implementirana unapređenja inicijalnog skupa podataka, koja mu povećavaju

vrednost i upotrebljivost u različitim tipovima aplikacija. Ceo proces unapređenja podeljen je na delove, pri čemu svaki deo pripada nekoj fazi u životnom ciklusu povezanih podataka (slika 1). Rezultujući skup podataka u kompresovanom Turtle formatu javno je dostupan za preuzimanje<sup>2</sup>, i njegova veličina iznosi preko 15GB, a takođe učitana je i u Virtuoso triplet-skladište, čiji je SPARQL interfejs javno dostupan<sup>3</sup>. Veličina njegove baze podataka iznosi 86GB. Sve transformacije prikazane u ovom poglavlju implementirane su *python* i *bash* skriptovima, javno dostupnim na GitHub-u<sup>4</sup>.

### 3.1. GeoSPARQL kompatibilnost

Generisani skup podataka nije GeoSPARQL kompatibilan, jer su geoprostorni podaci dati u formi GPS koordinata, a ne kako standard [20] nalaže geometrijskim entitetima *Feature* i *Geometry* koji su određeni WKT (engl. *well-known-text*) literalima. Ta mana onemogućuje aplikacije i sisteme za upravljanje RDF bazama podataka da koriste specifične upite koji sadrže geoprostorne predikate i funkcije, npr. *contains()*, *covers()*, *intersects()*, *touches()*, *distance()*, koje su preko potrebne u ovakvim scenarijima korišćenja [26]. Da bi se ovo prevazišlo i omogućila upotreba GeoSPARQL upita u aplikacijama koje koriste ove podatke, u fazi obogaćivanja i klasifikacije (slika 1), skup podataka je unapređen i usklađen sa standardom na način koji sledi. Transformacije putanja baziraju se na kreiranju WKT literala tipa *LINESTRING* za svaku generisanu putanju. Korišćeni generator je ograničen modelom podataka opisanim u prethodnom poglavlju, pa se transformacije podataka mogu sprovesti ili modifikacijama nad samim generatorom (primer takvih modifikacija opisan je u radu [27]), ili razvijanjem specifičnih transformacija nad već generisanim podacima, što i predstavlja pristup u ovom radu. Svaka putanja sada predstavlja i instancu entiteta *geo:Feature*, koja je u relaciji *geo:hasGeometry* sa entitetom *sf:LineString*. Ova dva entiteta su klase definisane GeoSPARQL standardom. Drugi od njih je u relaciji *geo:asWKT* sa WKT literalom koji predstavlja celu putanju (slika 2). Slične transformacije primenjene su i na tačke, uz adekvatne entitete *sf:Point* i WKT literalne tipa *POINT* definisane standardom (slika 2). Za razliku od ovih transformacija vezanih za GeoSPARQL standard, dodat je svakoj putanji numerički identifikator (predikat *numID*), relacija između putanje i njenih istaknutih tačaka, prve (*hasStartPoint*) i poslednje (*hasEndPoint*), kao i trajanje putovanja u sekundama (*hasDuration*). Deo skupa podataka kompatibilan sa GeoSPARQL standardom prikazan je na slici 2. Posle svih ovih poboljšanja, veličina skupa podataka se skoro udvostručila (preko 1532 miliona tripleta).

```

1 <#trace> a geo:Feature ;
2   numID      0^^xsd:integer ;
3   hasStartPoint  <#point0> ;
4   hasEndPoint  <#point0> ;
5   hasDuration  "1465"^^xsd:integer ;
6   geo:hasGeometry  <#trace_geom> ;
7   geo:asWKT    "LINESTRING(12.279363 51.322363,...,12.48659 51.481587)"^^geo:wktLiteral .
8
9
10 <#trace> :hasPoint  <#point0> .
11 <#point0> a geo:Feature ;:TracePoint ;
12   hasTimestamp  "2017-09-03T12:48:11Z"^^xsd:dateTime ;
13   lat  51.322363 ;
14   lon  12.279363 ;
15   hasSpeed  <#speed0> ;
16   geo:hasGeometry  <#point0_geom> .
17 <#point0_geom> a "POINT(12.279363 51.322363)"^^geo:wktLiteral .
18 <#speed0> a :velocityValue 30.00 ;
19   velocityMetric  kilometers_perHour .
20

```

Slika 2: Izvod iz GeoSPARQL kompatibilnog skupa podataka

### 3.2. Dodatna dimenzija podataka

Pored tačnih koordinata putanje nekog vozila, korisnicima ovog skupa podataka može od interesa biti i sama konfiguracija

<sup>2</sup> [http://download3.openlinksw.com/SAGE/datasets/Leipzig\\_Traffic\\_Data\\_5D.tar.gz](http://download3.openlinksw.com/SAGE/datasets/Leipzig_Traffic_Data_5D.tar.gz)

<sup>3</sup> <http://sage.demo.openlinksw.com/sparql>

<sup>4</sup> <https://github.com/mirkospasic/FloatingCarDataTransformation>

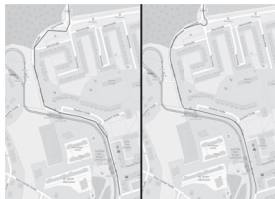
rute, tj. nagiba delova trase, da li je i koliko strm/ravan određeni deo putanje. Na primer, ova informacija može biti vrlo korisna biciklistima koji planiraju svoju rutu kroz planine, i slično. Imajući ovaj potencijal u vidu, skupu podataka su dodati podaci o nadmorskoj visini svih tačaka i putanja u njemu (faza ekstrakcije na slici 1). Ova dodatna merenja prikupljena su iz digitalnog elevacionog modela (DEM) Evrope, koji je rezultat projekta GMES RDA (EU-DEM) [28]. Model je javno dostupan na veb prezentaciji agencije Evropske unije za životnu sredinu (engl. *European Environment Agency*) u GeoTIFF formatu sa rezolucijom od 25m.

Za svaku tačku putanje, reprezentovanu vrednostima geografske širine i dužine, izvučena je njema nadmorska visina iz prethodno opisanog modela koristeći servis OpenElevation<sup>5</sup> instaliran uz pomoć Docker<sup>6</sup> sistema. Veb server opslužuje GET i POST zahteve, čiji su argumenti lokacije zadate geografskim koordinatama, a rezultat je njihova nadmorska visina u JSON formatu. Ovaj podatak je u skupu podataka pridružen svakoj tački svake putanje predikatom *elv*, i na taj način dodato je preko 126 miliona novih tripleta. Takođe, WKT literali tipa POINT (LINESTRING) postali su POINTZ (LINESTRINGZ) kako bi uključili i treću dimenziju u svoju reprezentaciju, shodno standardu.

### 3.3. Prilagođavanje mapi

Iako je ovaj skup podataka sintetički izgenerisan, on nije idealan u smislu preciznosti i sadrži probleme koji se sreću u realnosti. Na primer, GPS uređaji koji se koriste u vozilima za pozicioniranje i navigaciju ne daju savršene rezultate, i na preciznost njihovog merenja mogu uticati mnogi činioci, kao što su vremenske prilike, elektro-magnetno i solarno zračenje, itd. Dakle, GPS koordinate dobijene na ovaj način mogu odstupati od realne pozicije vozila, tj. šum u svakodnevnim merenjima je uvek prisutan. TomTom-ov generator oslikava takođe i ovu karakteristiku podataka, tako da aplikacije koje koriste ove podatke moraju da se izbore i sa time. Na primer, vizualizacija putanje (u navigacionom softveru, ili bilo kojoj drugog aplikaciji) koja većinu vremena prikazuje vozilo van puta uzrokuje loše korisničko iskustvo [29], što nije poželjno (slika 3).

Vizualizacija putanje vozila na mapi koja je često prikazana van puta, pored merenja koja sadrže šum, može biti prouzrokovana i intervalom vremena proteklim između dva uzastopna merenja. Na primer, u generisanim podacima on je 5 sekundi, što znači da je poznata lokacija vozila u trenutku  $t$ , i u trenutku  $t+5s$ , a o poziciji vozila između ta dva trenutka ne zna se ništa. Za to vreme, vozilo može da pređe i desetine metara, i da, na primer, u trenutku  $t$  bude ispred kružnog toka, a u trenutku  $t+5s$  bude iza njega. U tom vremenskom intervalu, reprezentacija putanje vozila bi bila duž koja bi prelazila preko sredine kružnog toka i/ili ne bi ispratila njegovu krivinu (slika 3). Ova detekcija problema spada u fazu analize kvaliteta (slika 1), koja je praćena fazom evolucije i popravke podataka, u koju i spada primenjen algoritam transformacije putanje opisan u tekstu koji sledi.



Slika 3: Vizualizacija putanje vozila pre (levo) i posle (desno) usklađivanja sa mapom

<sup>5</sup> <https://hub.docker.com/t/open-elevation/open-elevation/>

<sup>6</sup> <https://www.docker.com/>

Ovi problemi rešavaju se usklađivanjem sa mapom (engl. *map matching algorithms*), tj. procesom određivanja stvarne putanje vozila na mreži puteva na osnovu sekvence GPS opservacija koje sadrže šum. One moraju biti povezane sa ivicama postojeće mreže puteva u redosledu koji odgovara putanji vozila. Prema vremenu svog izvršavanja, ovi algoritmi dele se na algoritme u realnom vremenu (engl. *real-time, online, on the fly algorithms*) i one koji to nisu (engl. *offline algorithms*) [29]. Kod prvih, poravka GPS koordinata se radi neposredno, što je pre moguće, u trenutku njihovog prikupljanja, tj. kako se koja tačka sačuva, putanja do nje se popravljaju, bez bilo kakvih informacija o narednim tačkama putanje. Tipični scenariji gde se primenjuje ovaj tip algoritama su aplikacije satelitske navigacije (engl. *satellite navigation*). Za razliku od njih, kod druge vrste algoritama popravka putanje vrši se posle prikupljanja svih koordinata putanje i često se koriste u analizi transportnih sistema sa ogromnim količinama podataka. Glavna razlika između ova dva tipa je u performansama i preciznosti, i njihovom kompromisu koji je uvek prisutan. Prva vrsta algoritama mora da bude brza i efikasna, kako bi pružila odgovor korisniku u interaktivnom vremenu, ali je, sa druge strane, njena preciznost lošija, jer se algoritam izvršava nad manje poznatih tačaka. Zbog toga, drugi tip algoritama je pouzdaniji, ali se cena plaća u performansama. Kako je cilj ovog skupa podataka da se aplikacijama omogući analiza saobraćaja, predviđanje saobraćajnih gužvi, procena vremena putovanja, nad njim je primenjen drugonavedeni tip algoritama.

Za ovu svrhu, korišćena je GraphHopper<sup>7</sup> biblioteka otvorenog koda za rutiranje napisana u Javi. Ulaz u njen algoritam je niz izmerenih GPS koordinata, a izlaz takodje niz koordinata, ali onih koji odgovaraju realnoj putanji vozila koja se drži specificirane mreže puteva. Njen algoritam radi na sledeći način: za svaku tačku pronalazi koordinate kandidata na realnoj mreži puteva u odgovarajućem radijusu, a zatim Viterbi algoritmom [30] izabira najverovatniji niz uzastopnih kandidata, uzimajući u obzir rastojanje između izmerenih GPS koordinata i njihovih kandidata, kao i distancu između uzastopnih kandidata. Na primer, na slici 4, najjednostavniji algoritam bi svakoj izmerenoj GPS lokaciji pridružio tačku na najbližem putu, ali bi u tom slučaju druga i treća tačka bile pogrešno pridružene, tako da umesto uzimanja u obzir individualnih tačaka, mora se razmatrati njihova sekvenca.

Problem sa putanjom koju vraća GraphHopper algoritam je u nedostajućim informacijama pridruženim inicijalnim tačkama (vremena, brzina, nadmorska visina). Inicijalne tačke putanje najčešće nisu ni uključene u novu putanju, tj. algoritam ne obezbeđuje njihovu vezu sa novim segmentima putanje, pa je sledeći korak njihovo učešljavanje (engl. *merging*) u sekvencu novih tačaka. Preciznije, učešljavaju su projekcije inicijalnih tačaka na segment određen dvema novim tačkama između kojih se projekcija ubacuje, pa je time onemogućeno da bilo koja tačka nove putanje (i inicijalne i nove tačke) odstupa od idealne putanje vozila. Slika 3 sadrži vizualizaciju jedne od putanji iz skupa podataka pre popravke (levo), i posle (desno). Za razliku od inicijalnih tačaka, nove nemaju informaciju o vremenu, brzini i nadmorskoj visini kada se vozilo našlo na baš toj lokaciji. Nadmorska visina izračunata je na isti način kao za inicijalne tačke (iz EU-DEM modela), a vreme i brzina za svaku novu tačku dobijena je linearnom interpolacijom na osnovu vrednosti u dvema najbližim inicijalnim tačkama.

<sup>7</sup> <https://www.graphhopper.com/>



Slika 4: Usklađivane sa mapom (slika preuzeta iz [30])

### 3.4. Mreža segmenata puteva i veza sa LGD

Kako bi predviđanje mogućih saobraćajnih gužvi, uskih grla i trajanja putovanja bilo omogućeno, individualne putanje, njihova vremena i brzine nisu bitne, već je potrebno agregirati sve ove vrednosti na mrežu puteva, preciznije na njene najmanje segmente [29]. Za ovu svrhu, dodata je nova relacija (predikat *isLinked*) koja povezuje inicijalne tačke putanje, sa segmentima određenim dvema uzastopnim tačkama nove putanje između kojih je projekcija inicijalne tačke učešljana. Takođe, za te segmente uvedena je klasa *RoadSegment* kojoj oni pripadaju, i predikat *hasGeometry* koji ih povezuje sa geometrijama tipa *sf:LineString* koji dalje imaju WKT literale tipa *LINestring* određene tim dvema tačkama. Ovakva transformacija spada u fazu klasifikacije i obogaćivanja (slika 1).

Instance klase *RoadSegment* predstavljaju uvek segmente puta određene dvema tačkama, i za njih ne postoji nikakva informacija osim koordinata početne i krajnje tačke. Ako je neka dodatna informacija o njima potrebna (na primer da li se radi o delu auto-puta, magistrale, ili gradske ulice, da li je ulica jednosmerna, njeno ime, i ograničenje brzine) morao bi se dovesti u vezu sa putevima u nekom drugom skupu podataka koji sadrži sve potrebne informacije, recimo LGD skupu podataka, gde su oni reprezentovani klasom *HighwayThing* i njenim podklasama [24]. Njihovo nalaženje i materijalizacija korisnih informacija spadaju u fazu povezivanja i objedinjavanja (slika 1).

Jedan od problema pri ovakvom linkovanju je taj što segmenti puteva u LGD skupu podataka predstavljaju delove između dve raskrsnice, i ako postoji krivina na tom delu, oni nisu predstavljeni dvema tačkama, početnom i krajnjom, već i nekim između, tako da njihova geometrijska reprezentacija nije duž, tj. *LINestring* koji sadrži dve tačke, već izlomljena linija, tj. niz duži, *LINestring* sa više tačaka. U opštem slučaju, instancama klase *RoadSegment* odgovaraju pojedinačne duži od kojih se sastoji jedan segment puta u LGD skupu podataka. Odstupanje od ovoga mogu biti prvi i poslednji *RoadSegment* od kojih se sastoji putanja vozila, jer prva tačka odakle je vozilo krenulo ne mora biti raskrsnica, ili tačka na kojoj se *LINestring* LGD segmenta lomi, već može biti i neka tačka između. Sve ovo pravi dodatne probleme pri linkovanju instanci klase *RoadSegment* i LGD segmentata, pa nije dovoljno porediti geometrijske objekte na jednakost, već se mora proveriti da li su objekti u relaciji *covers*, tj. da li LGD segment prekriva instancu za koju se on traži. Kako se radi o jednodimenzionalnim objektima, provera ove relacije nije previše računarski zahtevna. Dodatni problem koji znatno komplikuje zadatak predstavlja različita preciznost podataka, tj. zaokruživanje koordinata u trenutnom skupu podataka i LGD-u, pa se relacija *covers* treba proveravati do na neku tačnost. U takvom scenariju se ne radi više o preklapanju jednodimenzionalnih objekata, već dvodimenzionalnih, tj. poligona, što je

računarski mnogo zahtevnije. Kako bi ovaj zadatak bio izvršen efikasnije, i kako ne bi bili računati preseki previše poligona, za svaki segment iz trenutnog skupa podataka se računa pripadajući pravougaonik koji ga sadrži (engl. *bounding box*), proširen na sve strane vrednostima greške zaokruživanja, i traže se LGD segmenti čiji pripadajući pravougaonik ima preseke sa već pomenutim pravougaonikom (njihov presek se vrlo jednostavno izračunava). Na taj način dobijaju se potencijalni kandidati, kojih može imati nekoliko, za koje se računa zahtevan presek poligona i određuje jedan od njih koji preklapa polazni segment. Poseban problem ovde je i određivanje preciznosti u kojoj se računa relacija *covers*. Prevelika preciznost može eliminisati sve kandidate, dok premala može ostaviti u igri više kandidata (na primer dva paralelna dela puta na maloj udaljenosti). Ovo je rešeno binarnom pretragom intervala željenih vrednosti preciznosti. Ako od potencijalnih kandidata ostaje jedan, pretraga se završava, ako je ostalo više, preciznost treba smanjiti i pokušati ponovo, a ako su svi eliminisani povećati je. Suština ovog linkovanja je sadržana u šablonu SPARQL upita prikazanom na slici 5, gde se parametri uokvireni simbolom @, tj. *seg1*, *seg1box*, *seg1wkt* i *prec*, redom menjaju konkretnim URI-jem polaznog segmenta, njegovim pripadajućim pravougaonikom, WKT literalom i željenom preciznošću. Umesto klase *HighwayThing* pri linkovanju korišćena je i klasa *RouteThing* [24], jer par desetina LGD segmenata predstavljaju putanje trajekta po obližnjim jezerima, a ne vozila po uobičajenim putevima. Na ovaj način svaki od preko 287 hiljada polaznih segmenata, povezan je (relacija *connected*) sa preko 30 hiljada LGD segmenata koji ih sadrže, a opcione vrednosti njihovih podklasa, labela, imena (lokalnih, internacionalnih, nacionalnih i kratkih) i geometrijski objekti koji ih opisuju biće materijalizovani i dodati u polazni skup podataka. Ovom izmenom, poslednjom u nizu, skup podataka je narastao na skoro 4 milijarde tripleta.

```

CONSTRUCT {
  @seg1 tontom:connected ?segment
  ?segment rdfs:label ?label
  ?segment rdfs:localName ?label
  ?segment rdfs:internationalName ?internationalName
  ?segment rdfs:nationalName ?nationalName
  ?segment rdfs:shortName ?shortName
  ?segment geom:geometry ?wkt
}

WHERE {
  ?type rdfs:subClassOf* lgdo:HighwayThing .
  ?segment a ?type .
  ?segment geom:geometry ?wkt .
  FILTER (bif:GeometryType(?wkt) = "LINestring")
  FILTER (bif:st:mayIntersect(?wkt, bif:st:geomFromText("@seg1box@")))
  FILTER (@seg1 rdfs:contains (geom:buffer(?wkt, @prec@), bif:st:geomFromText("@seg1wkt@")))
  OPTIONAL { ?segment rdfs:label ?label }
  OPTIONAL { ?segment lgdo:localName ?localName }
  OPTIONAL { ?segment lgdo:internationalName ?internationalName }
  OPTIONAL { ?segment lgdo:nationalName ?nationalName }
  OPTIONAL { ?segment lgdo:shortName ?shortName }
}
    
```

Slika 5: Upit za povezivanje LGD segmenata

### 4. PRIMERI UPOTREBE

Pripremljen skup podataka daje velike mogućnosti za pojednostavljen razvoj kompleksnih GIS aplikacija, koje se mogu koristiti svakodnevno sa ciljem analize, optimizacije i upravljanja saobraćajem, što je i demonstrirano u ovom poglavlju predstavljanjem scenarija korišćenja i primera upotrebe implementiranih u potpunosti u okviru funkcionalne GIS veb aplikacije<sup>8</sup>. Za pristup njoj, potrebna je autentifikacija<sup>9</sup>, pa se zainteresovani korisnici mogu javiti autoru rada elektronskom poštom. Tri modula GIS aplikacije, opisana u ovom poglavlju, demonstriraju faze kako skladištenja i upita nad podacima, tako i pretrage, vizualizacije i istraživanja nad njima (slika 1), ali je omogućen i način na koji bi se eventualne individualne greške u podacima mogle ispraviti (faza ručnih revizija). Takođe, aplikacija je jako korisna svakodnevno u oblastima analize, nadgledanja i upravljanja saobraćajem.

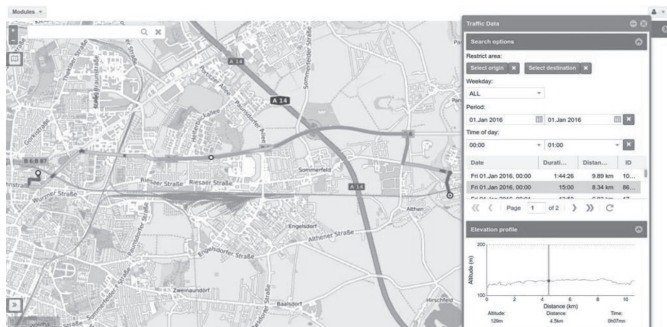
<sup>8</sup> <https://sage.geomer-maps.de/sage-shogun/login.html>

<sup>9</sup> Aplikacija je razvijana u sklopu šireg projekta, i kao takva sadrži druge nezavisne module koji sadrže poverljive podatke koji ne smeju biti javno dostupni.

Umesto skupa podataka koji se generiše sintetički, mogao bi se koristiti tok podataka (engl. *data stream*) prikupljen stvarnim GPS merenjima u realnom vremenu (u slučaju da je dostupan), i na potpuno isti način, istim transformacijama, imati aplikaciju koje radi sa realnim i trenutno aktuelnim podacima.

#### 4.1. Vizualizacija putanja vozila

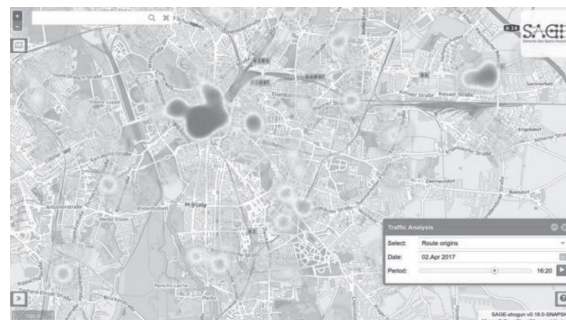
Implementirana aplikacija ima modul za filtriranje i prikaz putanji vozila na mapi. Filtriranje se može vršiti geoprostorno, po lokaciji početne i završne tačke putanje, kao i vremenski, zadavanjem vremenskog opsega od interesa, kako početkom i krajem intervala, tako i danima u nedelji. Željene putanje vozila predstavljaju se u listi, svojim identifikatorom, vremenom početka putovanja, ukupnim trajanjem i dužinom pređenog puta. Daljim izborom neke od njih, vidljivi deo mape biva prilagođen njoj (odgovarajuće izabran i zumiran), putanja biva iscrtana na mapi linijom čija boja može uzimati sve nijanse između zelene, preko žute i narandžaste do crvene, u zavisnosti od brzine vozila u odgovarajućem trenutku puta. Na slici 6, jasno se vidi različita brzina vozila bržim bulevarima i malim, uskim ulicama u gradskom jezgru. Takođe, evidentan je i pad brzine pri većim raskrnicama i ukrštanjima sa drugim ulicama. Izborom neke od putanja, dobija se vizualni prikaz njenog elevacionog profila (slika 6), preko koga kada se pređe mišem biva prikazana tačka na mapi koja odgovara tom delu putanje sa odgovarajućom nadmorskom visinom.



Slika 6: Modul za filtriranje i vizualizaciju putanja vozila u GIS veb aplikaciji

#### 4.2. Analiza saobraćaja

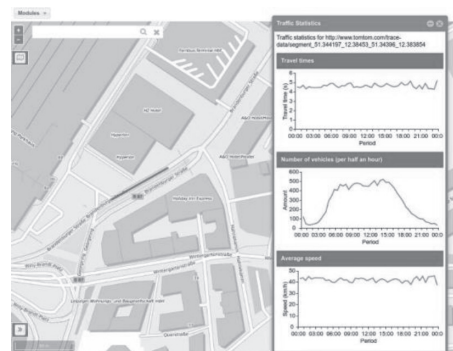
Drugi bitan modul aplikacije omogućava prikaz stanja na putevima u izabranom vremenskom intervalu u vidu toplotne mape (engl. *heat map*), tj. grafičke reprezentacije koja prikazuje stepen opterećenja saobraćaja različitim nijansama boja (slika 7). Toplotna mapa može biti generisana na osnovu svih aktivnih vozila u datom trenutku (i tada se očekivano može primetiti veći broj vozila u gradskom jezgru, nego na periferiji), ali takođe i samo onih koji tada počinju ili završavaju putovanje, pa se da zaključiti za izabrani trenutak odakle većina stanovnika otpočinje, a gde završava svoje putovanje. Na primer, radnim danima u jutarnjim časovima evidentan je veći broj početnih tačaka putanja u delovima grada sa stambenim blokovima, a krajnjih u centralnoj četvrti gde se nalazi većina poslovnih prostora i kancelarija, dok je u popodnevnom časovima obrnuto. Aplikacija nudi i klizač (engl. *slider*), koji se u izabranom vremenskom intervalu pomera automatski od početka zadatog intervala unapred, prikazujući vreme za koje je prikazana toplotna mapa generisana, i tako na mapi možemo videti interaktivnu simulaciju, tj. glatko pomeranje gustine saobraćaja iz jedne oblasti u neku drugu (slika 7).



Slika 7: Modul za prikaz opterećenosti saobraćaja u GIS veb aplikaciji

#### 4.3. Statistike saobraćaja

Pomenuta aplikacija sadrži i modul za prikaz različitih statistika saobraćaja koje su agregirane na celoj mreži segmentata puteva. Dozvoljen je izbor jednog ili više segmentata na mapi, kojim se otvara grafikon prikaza prosečnog vremena putovanja na tom delu puta po časovima u toku dana, broj vozila koja su prošla izabranom rutom distribuiranih po vremenskoj osi u toku dana, kao i njihove prosečne brzine. Na slici 8, jasno se vidi da za izabrani segment nema većih oscilacija u brzini i vremenu potrebnom za prelazak, što ukazuje da on nije usko grlo saobraćaja na kome se prave gužve u saobraćajnim špičevima (engl. *rush hour*). Prosečan broj vozila koja prolaze izabranim segmentom najveći je ujutru oko 8 i posle podne oko 15 časova, kao i između, dok je van tog vremenskog intervala osetno niži (slika 8). Ovaj modul može se lako dopuniti izborom dana u nedelji (meseca u godini) čiji saobraćaj ulazi u statistike, i na osnovu toga analizirati uticaj radnih i neradnih dana (letnjih meseci) na povećan/smanjen saobraćaj.



Slika 8: Modul za prikaz statistika segmentata mreže puteva u GIS veb aplikaciji

#### 4.4. Predikcija vremena putovanja

Prikazani statistički modul aplikacije iz potpoglavlja 4.3 može se lako unaprediti tako da prikazuje prediktivno trajanje putovanja između izabranih tački A i B na mapi, u zavisnosti od željenog vremena početka putovanja. Agregacijom već izračunatih potrebnih vremena za prolazak svakog segmenta na putu od tačke A do tačke B ponaosob, u odgovarajućem periodu dana, može se dobiti grafikon ukupnog vremena putovanja u toku celog dana. Ovakva predikcija može biti od velike koristi osobama koje svoje putovanje mogu pomeriti za malo ranije ili kasnije od željenog, i na taj način znatno im uštedeti vreme izbegavanjem gužvi. Korisnike koji imaju fiksirano vreme početka putovanja aplikacija može informisati i upozoriti na procenjeno duže trajanje putovanja od uobičajenog usled povećane saobraćajne aktivnosti.

## 5. ZAKLJUČAK

U ovom radu prikazana je evolucija skupa podataka vozila u pokretu kroz ceo životni ciklus povezanih podataka, prateći sva načela njihovih transformacija i objavljivanja. Počevši od podataka softverski generisanih, koji su sadržali samo geolozijske informacije (GPS koordinate) i brzinu vozila u kratkim vremenskim intervalima, konstruisan je višedimenzionalni skup podataka, kompatibilan sa GeoSPARQL standardom, u kome su putanje vozila prilagođene mreži puteva, i povezane sa ogovarajućim segmentima i korisnim informacijama u LGD skupu podataka, koji predstavlja glavnu referentnu tačku u sferi geoprostornih povezanih podataka. Evolucija je prikazana korak po korak, i za svaki od njih identifikovana je odgovarajuća faza u životnom ciklusu. Finalni skup podataka, meren u desetinama gigabajta i milijardama tripleta, javno je dostupan, spreman da bude pohranjen u neko od triplet-skladišta sa podrškom za geoprostorne podatke. Takođe, podaci su učitani u Virtuoso, i može im se pristupiti preko javno dostupnog SPARQL interfejsa, kako interaktivno, tako i upitima iz aplikacija različitih domena povezanih sa vizualizacijom geoprostornih podataka ili analize, nadgledanja i upravljanja saobraćajem sa ciljem identifikacije uskih grla, saobraćajnih gužvi i sličnih problema. Prikazani su i scenariji korišćenja skupa podataka, u potpunosti razvijeni u okviru tri različita modula prikazane GIS veb aplikacije, koji demonstriraju značaj i mogućnosti pripremljenih podataka, kao i važnost njihove transformacije u skladu sa standardima i principima javno dostupnih povezanih podataka.

Dalja unapređenja uključuju učitavanje podataka u triplet-skladište ne odjednom, već simulirajući realni tok podataka, tj. učitavajući putanje u skladu sa vremenima njihovih poslednjih tačaka, pri čemu je moguće zadati parametar koji će definisati kompresiju vremena simulacije u odnosu na realno vreme, i na taj način simulirati saobraćaj iz vremenskog intervala od dve godine za par sati, dana, ili nedelja. Upiti nad trenutnim stanjem baze, uzimajući u obzir samo novopridošle podatke, postaću kontinualni upiti (engl. *continuous queries*) i odražavaće stanje saobraćaja u realnom vremenu. Na ovaj način, razvijene aplikacije biće spremne za tranziciju sa generisanih podataka, na stvarne, koji će biti prikupljeni u realnom vremenu, transformisani na način opisan u ovom radu, i odmah potom sačuvani u triplet-skladištima spremni za fazu eksploatacije.

**ZAHVALNICA:** Ovaj rad je finansiran od strane Evropske komisije (engl. *European Commission*) kroz Eurostars projekat SAGE (engl. *Semantic Geo-Spatial Analytics*), zaveden pod brojem GA no. E!10882.

## 6. LITERATURA

- [1] Patel, J. (2019). *Bridging Data Silos Using Big Data Integration*, International Journal of Database Management Systems 11 : 01-06.
- [2] Ziegler, P. and Dittrich, K. R. (2007). *Data Integration — Problems, Approaches, and Perspectives*. In: (Ed.), *Conceptual Modelling in Information Systems Engineering*, Springer Berlin Heidelberg.
- [3] Hitzler, P.; Krotzsch, M. and Rudolph, S., (2009). *Foundations of Semantic Web Technologies*. Chapman and Hall/CRC.
- [4] Kashyap, V.; Bussler, C. and Moran, M., (2008). *The Semantic Web*. Springer Berlin Heidelberg.
- [5] Bizer, C.; Heath, T. and Berners-Lee, T. (2009). *Linked Data - The Story So Far*, International Journal on Semantic Web and Information Systems 5 : 1-22.

- [6] Auer, S.; Lehmann, J. and Ngomo, A.-C. N. (2011). *Introduction to Linked Data and Its Lifecycle on the Web*. In: (Ed.), *Reasoning Web. Semantic Technologies for the Web of Data*, Springer Berlin Heidelberg.
- [7] McBride, B. (2002). *Jena: a semantic Web toolkit*, IEEE Internet Computing 6 : 55-59.
- [8] Nenov, Y.; Piro, R.; Motik, B.; Horrocks, I.; Wu, Z. and Banerjee, J. (2015). *RDFox: A Highly-Scalable RDF Store*. In: (Ed.), *The Semantic Web - ISWC 2015*, Springer International Publishing.
- [9] Potocki, A.; Polukhin, A.; Drobyazko, G.; Hladky, D.; Klintsov, V. and Unbehauen, J. (2013). *OntoQuad: Native High-Speed RDF DBMS for Semantic Web*. In: (Ed.), *Knowledge Engineering and the Semantic Web*, Springer Berlin Heidelberg.
- [10] Erling, O. and Mikhailov, I. (2009). *RDF Support in the Virtuoso DBMS*. In: (Ed.), *Studies in Computational Intelligence*, Springer Berlin Heidelberg.
- [11] Güting, R. H. (1994). *GraphDB: Modeling and Querying Graphs in Databases*. In: (Ed.), *VLDB '94, Proceedings of 20th International Conference on Very Large Data Bases*, Morgan Kaufmann : 297-308.
- [12] Lal, M., (2015). *Neo4j graph data modeling*. Packt Publishing Ltd.
- [13] Prud'hommeaux, E. and Seaborne, A. (2008). *SPARQL Query Language for RDF, W3C Recommendation*, <http://www.w3.org/TR/rdf-sparql-query/>, datum pristupa 23.04.2020.
- [14] Baralis, E.; Valle, A. D.; Garza, P.; Rossi, C. and Scullino, F. (2017). *SQL versus NoSQL databases for geospatial applications*, IEEE International Conference on Big Data, IEEE.
- [15] TomTom International, BV (2020). *About Us*, <https://www.tomtom.com/company/>, datum pristupa 23.04.2020.
- [16] Zdravković, S. and Minović, M. (2016). *Primeni i pravci budućeg razvoja bežične komunikacije u pametnim transportnim sistemima*, Info M 15 : 18-22.
- [17] OpenStreetMap Contributors (2015). *Planet dump*, <https://planet.openstreetmap.org>, datup pristupa 23.04.2020.
- [18] Ioannidis, T.; Garbis, G.; Kyzirakos, K.; Bereta, K. and Koubarakis, M. (2019). *Evaluating Geospatial RDF stores Using the Benchmark Geographica 2*, ArXiv.
- [19] Battle, R. and Kolas, D. (2012). *Enabling the geospatial Semantic Web with Parliament and GeoSPARQL*, Semantic Web 3 : 355-370.
- [20] Open Geospatial Consortium (2012). *GeoSPARQL - A Geographic Query Language for RDF Data*, <https://www.opengeospatial.org/standards/geosparql>, datum pristupa 23.04.2020.
- [21] Spasić, M. (2015). *Design of Geospatial Benchmarking System and Performance Evaluation of Virtuoso and PostGIS*. In: (Ed.), *ICIST 2015 Proceedings 1* : 154-159.
- [22] Huang, W.; Raza, S. A.; Mirzov, O. and Harrie, L. (2019). *Assessment and Benchmarking of Spatially Enabled RDF Stores for the Next Generation of Spatial Data Infrastructure*, ISPRS International Journal of Geo-Information 8 : 310.
- [23] Jovanović, M.; Williams, H. and Spasić, M. (2019). *Prototype Verifying Virtuoso GeoSPARQL and DE-9IM Compliance, Project deliverable from Project SAGE*.
- [24] Stadler, C.; Lehmann, J.; Höffner, K. and Auer, S. (2012). *LinkedGeoData: A core for a web of spatial open data*, Semantic Web 3 : 333-354.
- [25] Bösche, K.; Sellam, T.; Pirk, H.; Beier, R.; Mieth, P. and Manegold, S. (2013). *Scalable Generation of Synthetic GPS Traces with Real-Life Data Characteristics*. In: (Ed.), *Selected Topics in Performance Evaluation and Benchmarking*, Springer Berlin Heidelberg.
- [26] Jovanović, M. and Spasić, M. (2019). *Transforming Geospatial RDF Data into GeoSPARQL-Compliant Data: A Case of Traffic Data*.
- [27] Spasić, M.; Jovanović, M. and Prat-Pérez, A. (2016). *An RDF Dataset Generator for the Social Network Benchmark with Real-World Coherence*. In: (Ed.), *Proceedings of the Workshop on Benchmarking Linked Data co-located with the 15th International Semantic Web Conference (ISWC) 1700*.
- [28] Spasić, M.; Jovanović, M. and Williams, H. (2019). *Prototype Verifying 5D Geospatial Support in Virtuoso. Project deliverable from Project SAGE*.
- [29] Spasić, M.; Jovanović, M. and Williams, H. (2019). *Continuous Query Optimisations in GeoSPARQL. Project deliverable from Project SAGE*.
- [30] Newson, P. and Krumm, J. (2009). *Hidden Markov map matching through noise and sparseness*. In: (Ed.), *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM Press.



**Mirko Spasić**, asistent na Katedri za računarstvo i informatiku  
**Kontakt:** mirko@matf.bg.ac.rs  
**Oblasti interesovanja:** Verifikacija softvera, Upitni jezici, Povezani podaci

CIP – Katalogizacija u publikaciji Narodna biblioteka Srbije, Beograd 659.25:004

**INFO M** : časopis za informacione tehnologije i multimedijalne sisteme = journal of information technology and multimedia systems / glavni i odgovorni urednik Miroslav Minović. - [Štampano izd.] - God. 1, br. 1

(2002)- . - Beograd : Fakultet organizacionih nauka, 2002- (Smederevo : Newpress). - 30 cm

Dva puta godišnje. - Je nastavak: Info Science = ISSN 1450-6254. - Drugo izdanje na drugom medijumu: Info M (Online) = ISSN 2683-3646

ISSN 1451-4397 = Info M (Štampano izd.)

COBISS.SR-ID 105690636