

Mogućnosti besplatne verzije Rasdaman softvera za analize prostorno-vremenskih rasterskih podataka

OGNJEN S. ANTONIJEVIĆ, Univerzitet u Beogradu,
Građevinski fakultet, Beograd

Stručni rad
UDC: 629.786:004.651
DOI: 10.5937/tehnika1902198A

Snimci satelitskih misija su važan izvor rasterskih podataka sa mnogobrojnim oblastima primene u nauci i privredi. Istorijski podaci zajedno sa rastućom arhivom snimaka aktivnih misija predstavljaju velike količine podataka. Sentinel-2 misija lansirana 2015. godine je poslednja u nizu satelitskih misija čiji je cilj globalno prikupljanje podataka visoke rezolucije sa visokom vremenskim frekvencijom. Podaci ove misije su besplatno dostupni, za naučne i komercijalne svrhe. Izazov je te podatke analizirati na efikasan način zbog njihovog obima – prema izveštaju za 2017. godinu [1], dnevno se generiše 10TB satelitskih podataka, dok je veličina svih snimaka premašila 40PB. Rasterske baze podataka nude rešenje za brzo, efikasno i pouzdano manipulisanje velikim količinama rasterskih podataka. U ovom radu su testirane mogućnosti rasdaman-a, korišćenjem vremenske serije Sentinel-2 snimaka nad test područjem.

Ključne reči: vremenska serija, Sentinel-2, rasterske baze podataka

1. UVOD

Osmatranje i praćenje globalnih prirodnih fenomena značajno je za čovečanstvo. Satelitski snimci Zemlje su važan izvor podataka, sa dugom istorijom i brojnim oblastima primene. Vremenske serije ovih podataka daju mogućnost praćenja fenomena kroz vreme. Sentinel misija je najnovija u nizu satelitskih misija koje omogućavaju osmatranje Zemljišnog pokrivača kroz prostor i vreme.

Analiza vremenskih serija satelitskih snimaka zahteva odgovarajući softver i hardver. Rasterske baze podataka [2] pružaju mogućnosti za efikasno manipulisanje velikim količinama rasterskih podataka – one imaju sve prednosti relacionih baza podataka (konzistentnost, sigurnost, optimizacija upita, distribuiranost...). Jedan od pionira na tom polju je i rasdaman (raster data manager) [3].

Cilj rada je da istraži mogućnosti rasdaman-a za analize prostorno-vremenskih rasterskih podataka. Korišćena je vremenska serija Sentinel-2 snimaka, koji predstavljaju promenu NDVI indeksa kroz vreme na test području. NDVI (Normalized Difference Vegetati-

on Index) indeks predstavlja normalizovanu razliku između reflektovane svetlosti u infracrvenom (koji vegetacija skoro potpuno reflektuje) i crvenom (koji vegetacija absorbuje) delu spektra. Raspon vrednosti indeksa je od -1 do 1. Visoke vrednosti su indikator zdrave i zelene vegetacije. Ovaj indeks je povezan sa sadržajem hlorofila u vegetaciji, i ima značajnu primenu u poljoprivredi, šumarstvu, i drugim naukama [14]. Rasdaman je dostupan u komercijalnoj, i besplatnoj (community) verziji koja je korišćena u ovom radu.

2. TEHNOLOGIJE

Rasdaman je rasterska baza podataka koja podržava čuvanje, manipulaciju, i upite nad masivnim multidimenzionalnim nizovima. Rasteri se u rasdaman-u tretiraju kao višedimenzionalni nizovi vrednosti. Kod unosa u bazu raster se po zadatoj šemi [4], [5] deli na manje rastere (tajlove), što omogućava efikasniji pristup podacima (slika 1). Tajlovanjem je omogućena optimizacija prema očekivanim upitima korisnicima (podaci se dele tako da im korisnici najbrže pristupaju).

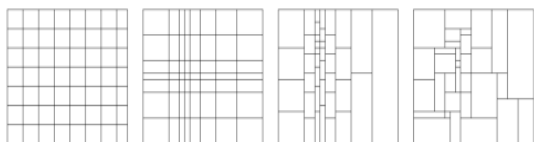
Integracija rastera u klasične baze podataka omogućena je čuvanjem tajlova u tabeli kao binarnih objekata. Uz svaki raster čuva se prostorni indeks, koji sadrži informacije o tajlovima rastera i njihovom prostornom rasporedu. Tako se prilikom izvršavanja upita samo potrebni delovi rastera učitavaju u memoriju.

Adresa autora: Ognjen Antonijević, Univerzitet u Beogradu, Građevinski fakultet, Beograd, Bulevar kralja Aleksandra 73

e-mail: antonijevic.ognjen@gmail.com

Rad primljen: 22.01.2019.

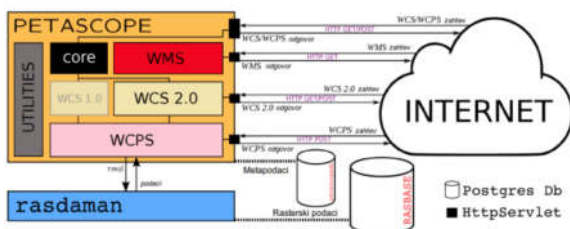
Rad prihvaćen: 14.03.2019.



Slika 1 – Primeri regularnog i neregularnog tajlovanja podataka

Strukture podataka u bazi i upiti nad podacima se definišu preko posebnog jezika za upite (rasql-rasdaman query language). U bazi se raster definiše sa dva parametra: tipom vrednosti u pikselima rastera i brojem piksela po svakoj dimenziji – podacima o tipu i domenu rastera.

Petascopie [6] je web komponenta rasdaman-a koja implementira WMS [7], WCS [8], WCS-T [8], WCPS [9] standarde za serviranje rasterskih podataka. Ova komponenta je Java web servlet aplikacija koja radi na Tomcat-u. Rasteri sa geoprostornim podacima u bazu se unose i serviraju korisnicima kroz ovaj servis. Pri unosu, podaci rastera se upisuju i čuvaju u rasdaman-u, dok se metapodaci poput georeference čuvaju u posebnoj Postgresql bazi (slika 2).



Slika 2 – Arhitektura petascopie komponente

Rasteri se u petascopie unose preko WCS-T zahteva, korišćenjem skripta za učitavanje rastera i odgovarajućeg konfiguracionog fajla. Ovaj skript se poziva iz komandne linije, sa putanjom ka konfiguracionom fajlu. Skript automatski generiše za svaki raster URL zahteve za unos podataka, koje onda šalje servisu. Korisnik u konfiguracioni fajl upisuje potrebna podešavanja – lokaciju rastera koji se unose, url servisa, tip i naziv kolekcije u koju se smeštaju podaci, shemu tajlovanja, i ostalo podešavanja. Za vremenske podatke postoje dva predefinisana konfiguraciona fajla – za regularnu (sa jednakim vremenskim razmacima između snimaka) i neregularnu vremensku seriju.

Petascopie ima i web interfejs koji korisnicima omogućava da lako pregledaju metapodatke o dostupnim rasterima, da preuzimaju delove ili cele rastere i da formiraju upite i pregledaju rezultate u interaktivnoj konzoli.

3. PODACI

Sentinel-2 [10] satelitska misija je razvijena od strane Evropske svemirske agencije kao deo Kopenikus [11] programa. Čine je dva satelita sa 13 kanala u vidljivom, blisko infracrvenom i kratko-talasnom

infracrvenom delu elektromagnetnog spektra. Kanali su prostorne rezolucije 10, 20 i 60 m. Vremenska rezolucija podataka je 5 dana (za isti ugao snimanja) u najgorem slučaju, često i manje zbog preklopa između linija prolaza satelita. Podaci misije su javno dostupni i besplatni za korišćenje u naučne i komercijalne svrhe. Snimci su servirani po tajlovima – površina Zemlje je podeljena na grid sa ćelijama (tajlovima) dimenzije 100x100 km.

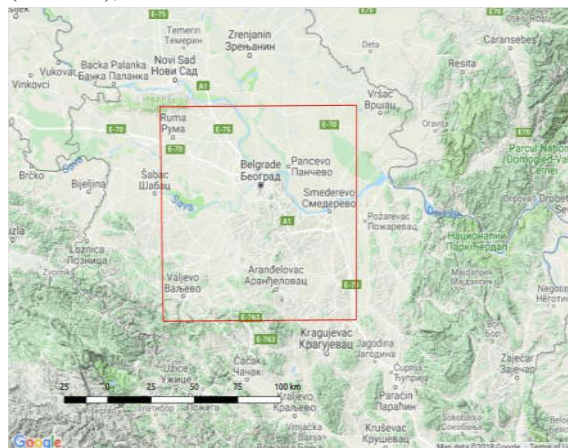
Predviđene primene podataka ove misije su između ostalih i praćenje vegetacije i ocena biofizičkih parametara. NDVI (Normalized Difference Vegetation Index) se dugi niz godina koristi za praćenje vegetacije. Računa se iz reflektovanog zračenja u crvenom i blisko-infracrvenom delu spektra:

$$NDVI = \frac{(R_{NIR} - R_{RED})}{(R_{NIR} + R_{RED})}$$

Ovaj indeks je korelisan sa sadržajem hlorofila u vegetaciji, samim tim i sa zdravljem vegetacije, biomasom, LAI (Leaf Area Index) indeksom, delom absorbovane radijacije (FAPAR indeks).

4. EKSPERIMENT

Za potrebe eksperimenta preuzeti su snimci jednog Sentinel-2 tajla za vremenski period od 16 meseci (april 2017. – septembar 2018.). Odabran je tajl 34-TDQ koji obuhvata područje centralne Srbije i deo Vojvodine (slika 1). Snimci su atmosferski korigovani korišćenjem Sen2Cor [[13]] alata. Delovi snimaka koji su pod oblacima su maskirani korišćenjem rastera sa podacim o oblačnosti, dobijenih iz procesa atmosferske korekcije. Iz maskiranih snimaka sračunat je NDVI za svaki datum. Test podaci obuhvataju 133 datuma (snimaka), veličine 37Gb.



Slika 3 – Test područje

Pripremljeni podaci ubačeni su u *rasdaman* preko *petascopie* servisa, u neregularnu vremensku kolekciju podataka. Po uspešnom unosu, podaci postaju dostupni na web servisu, gde je moguće njihovo potpuno ili

delimično preuzimanje u različitim formatima koristeći WCS GetCoverage zahtev.

Moguće je izdvojiti samo deo podataka zadavanjem prostornih i vremenskih intervala, biranjem specifičnih kanala, ili navođenjem granica poligona od interesa. Takođe je moguće reprojektovati i skalirati podatke.

Primer GetCoverage zahteva:

```
http://localhost:8080/rasdaman/ows?&
SERVICE=WCS&VERSION=2.0.1&
REQUEST=GetCoverage&
COVERAGEID=ndvi_timeseries&
SUBSET=E(430221,436788)&
SUBSET=N(4921030,4923560)&
SUBSET=ansi("2018-09-04T00:00:00.000Z")&
FORMAT=image/tiff
```

WCPS upiti omogućavaju procesiranje podataka na serveru, i pružaju veće mogućnosti od GetCoverage zahteva. Korisnici upite mogu slati preko URL zahteva u sklopu ProcessCoverages servisa, ili interaktivno preko web konzole.

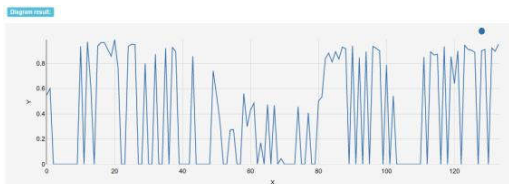
Jezik za pisanje ovih upita [9], [12] podržava različite operacije nad pikselima: logičke operacije, poređenja, trigonometrijske, logaritamske i kvadratne funkcije, transformacije tipova podataka, operacije agregacije (minimum, maksimum, srednja vrednost). Kombinacijom ovih operacija korisnici mogu praviti upite proizvoljne složenosti.

Najjednostavniji primer upita je praćenje promene vrednosti NDVI kroz vreme za lokaciju:

```
diagram>>for i in (ndvi_timeseries) return encode (
  i[E(400001),N(4894000),ansi("2017-04-
01":"2018-09-04")])
, "csv")
```

Rezultat je niz vrednosti koji je moguće dobiti u gml, csv ili json formatu. Moguće je i prikazivanje vrednosti u konzoli putem dijagrama (slika 4), dodavanjem "diagram>>" na početak upita. Prilikom testiranja ove funkcionalnosti uočeno je da program ima problem sa Nan vrednostima.

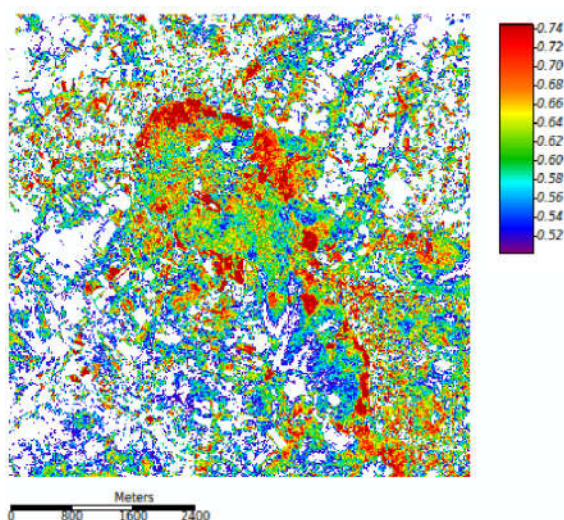
Kada su ove vrednosti prisutne u rezultatu upita, dijagram se ne prikazuje. Sa druge strane, kada se umesto ovih vrednosti stave nule, dijagram gubi na jasnoći prikaza.



Slika 4 – Promena NDVI indeksa kroz vreme u pikselu

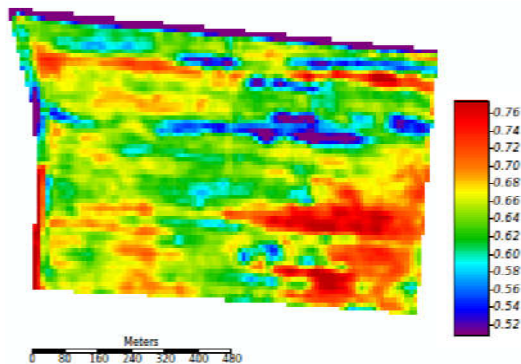
Podatke je moguće filtrirati po zadatom kriterijumu. Tako je moguće izdvojiti piksele čija je vrednost veća od zadate vrednosti. Sledeći upit na području oko Avale za 11.11.2017. izdvaja piksele čiji je NDVI>0.5 (slika 5):

```
for i in (ndvi_timeseries) return encode(
i[E(402000:408000),N(4985000:4989000),ansi("2017-
11-
12")]*(i[E(400000:401000),N(4985000:4986000),
ansi("2017- 07-11")]>0.6)
,"tiff")
```



Slika 5 – Područje Avale, pikseli sa NDVI>0.5. Na slici se jasno izdvajaju područja pod šumom

Dalji upiti neće biti prikazani u radu zbog obima teksta. Kod kompleksnijih upita postaje teško jasno pisanje i povećava se mogućnost greške, zbog potrebe za ponavljanjem operacije izdvajanja („odsecanja“) delova rastera – $i[E(x1:x2), N(y1:y2), ansi(t1:t2)]$, gde su $x1...t2$ prostorne i vremenske granice dela koji se izdvaja iz kolekcije. Primera radi, kod računanja NDVI za određeni datum i područje, potrebno je četiri puta ponoviti operaciju „odsecanja“ rastera.

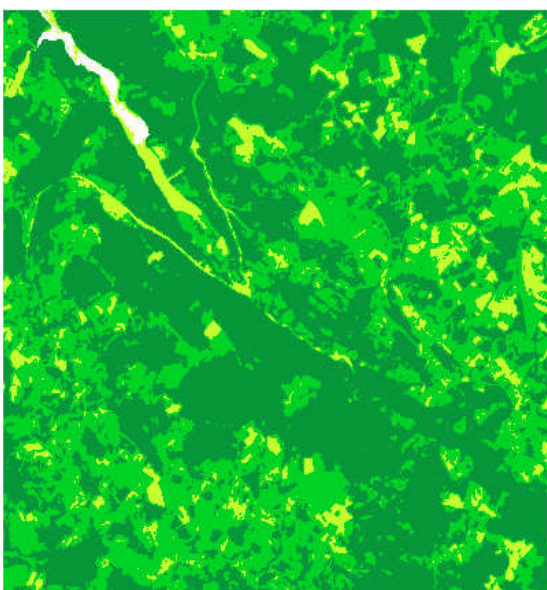


Slika 6 – Upit po granicama parcele

Korišćenjem funkcije clip moguće je izdvojiti piksele rastera po poligonu (slika 6). Poligon se zadaje u WKT formatu, sa url adresom definicije koordinatnog sistema u kome je zadat (npr. <http://open-gis.net/def/crs/EPSSG/0/4326>).

Još jedna korisna mogućnost jeste pravljenje karata "u letu" prema korisnički zadatim parametrima. Različiti rasponi vrednosti mogu se prikazati proizvoljnim bojama (slika 7). U navedenom primeru su vrednosti NDVI < 0.4 prikazane svetlo zelenom bojom, 0.4-0.8 zelenom i vrednosti > 0.8 tamno zelenom bojom. Ostali pikseli prikazani su belom bojom. Ova funkcionalnost pruža korisniku mogućnost za interaktivnu analizu podataka direktno u web konzoli, tako što se na početak upita doda „image>>“. Boje se zadaju u obliku {red: x; green: x; blue: x}.

2D Image result (png, jpeg):



Slika 7 – Dodeljivanje boja pikselima na osnovu zadatih vrednosti

Zbog obima i kompleksnosti upita, dešava se da korisnik napravi sintaksnu ili semantičku grešku. U ovim slučajevima, grešku je najlakše naći i otkloniti uvidom u logove Tomcat i rasdaman aplikacija. Logovi se nalaze u folderima /var/log/tomcat8 i /opt/rasdaman/log.

Prilikom testiranja uočeni su i nedostaci programa. Jedan od većih problema za korisnike predstavlja nemogućnost dodavanja snimaka između već postojećih datuma u kolekciji. Nov snimak je moguće dodati samo ako je njegov datum posle datuma poslednjeg ubačenog snimka. U praksi je ovo veliki problem, jer se često dešava da podaci postaju dostupni različitim brzinom, sa zakašnjenjem. Ako se novi podaci odnose na vremenski period pre poslednjeg podatka u bazi,

potrebno je ponovo uneti sve podatke u novu kolekciju.

Još jedan uočen problem jeste dimenzija ulaznog rastera kod učitavanja. Jednostavnim upitom za izdavanje dela rastera za proizvoljni datum, empirijski je utvrđeno da je približan limit u broju piksela rastera koji se unosi 70000x30500.

Pri povećanju rastera za 500 piksela bio kojoj dimenziji, program ne uspeva da unese raster u bazu, i prekida rad sa nedokumentovanom greškom. Testiranjem različitih dimenzija, utvrđeno je da limit zavisi od ukupnog broja piksela u rasteru, nezavisno od broja piksela po dimenziji.

Na sličan način je utvrđeno i ograničenje programa u pogledu veličine ulaznog rastera na disku. Menjanjem tipa podataka ćelija rastera (Byte, Int32, Int64, Float32, Float64), utvrđeno je da je limit u veličini ulaznog fajla ~ 4.5Gb.

5. ZAKLJUČAK

Rasdaman nudi celokupno rešenje za serviranje geoprostornih rasterskih podataka, i implementira sve relevantne OGC standarde. Podržan je veliki broj formata podataka. Projekat je dobro dokumentovan, sa dostupnim materijalima i tutorijalima za korišćenje.

Sa druge strane, korišćenje programa zahteva od korisnika poznavanje programiranja i rad u komandnoj liniji, što ne odgovara svim korisnicima.

Testiranje je pokazalo da rasdaman nije potpuno skalabilan, postoje ograničenja u dimenzijama i veličini rastera koji je moguće uneti u bazu. Takođe, nedostaje fleksibilnost pri dodavanju novih podataka, zbog nemogućnosti učitavanja novih snimaka između već postojećih datuma u bazi. U praktičnim primenama, ovo može biti veliki nedostatak.

Korisnicima su na raspolaganju različite mogućnosti manipulacije podacima. WCPS dozvoljava formiranje upita proizvoljne složenosti, i podržava sve važnije aritmetičke i druge operacije nad pikselima i rasterima. Međutim, kompleksnost jezika i veličina upita u praksi može predstavljati problem. Veći upiti su nečitljivi a jezik za upite često nije dovoljno deklarativan. Performanse programa su limitirane ograničenjem u broju procesorskih jezgara koja se koriste pri izvršavanju upita – besplatna verzija programa koristi jedno jezgro po upitu, što je značajno ograničenje.

Rasdaman je jedina implementacija WCPS standarda, ni jedan drugi program na tržištu ne objedinjuje sve funkcionalnosti koje su podržane u rasdaman-u. Uprkos navedenim nedostacima, razvoj programa se kreće u dobrom pravcu. Problem koji rasdaman pokušava da reši je aktuelan i značaj i uloga rasterskih baza podataka će rasti u budućnosti.

LITERATURA

- [1] Sentinel Data Access Annual Report 2017, preuzeto 24.9.2018. sa sentinel.esa.int/
- [2] Liu, Ling, Özsu, M. Tamer (Eds.) *Encyclopedia of Database Systems*, Springer, 2009
- [3] Rasdaman - raster data manager, DOI: 10.5281/zenodo.1040170
- [4] P. Furtado, P. Baumann *Storage of Multidimensional Arrays based on Arbitrary Tiling*. ICDE'99, March 23-26, Sydney, Australia, 1999.
- [5] Tiling strategies, preuzeto 24.9.2018 sa rasdaman.org/wiki/Tiling
- [6] A. Aiordachioaie, P. Baumann, *PetaScope: An Open-Source Implementation of the OGC WCS Geo Service Standards Suite*. Proc. Intl. Conf. on Scientific and Statistical Database Management (SSDBM'2010), Heidelberg, Germany, June 30 - July 2, 2010.
- [7] Web Map Service (WMS), preuzeto 24.9.2018. sa openeospatial.org/standards/wms
- [8] Web Coverage Service (WCS), preuzeto sa openeospatial.org/standards/wcs 20.9.2018.
- [9] Web Coverage Processing Service (WCPS), preuzeto sa openeospatial.org/standards/wcps 19.9.2018.
- [10] Sentinel-2 Mission, preuzeto 10.9.2018. sa sentinel.esa.int/web/sentinel/missions/sentinel-2
- [11] Copernicus program, preuzeto 12.9.2018. sa www.copernicus.eu/
- [12] P. Baumann: The OGC Web Coverage Processing Service (WCPS) Standard. *Geoinformatica*, 14(4)2010, pp 447-479
- [13] Sen2Cor - processor for Sentinel-2 Level 2A product generation and formatting, preuzeto 12.3.2019. sa <http://step.esa.int/main/third-party-plugins-2/sen2cor/>
- [14] Pettorelli N, Vik J. O, Mysterud A, Gaillard J. M, Tucker C. J. & Stenseth N. C, *Using the satellite-derived NDVI to assess ecological responses to environmental change*. *Trends in Ecology and Evolution*, 20(9), 503–510. <https://doi.org/10.1016/-j.tree.-2005.05.011>, 2005.

SUMMARY

RASDAMAN AS A PLATFORM FOR SERVING SPACE-TIME RASTER DATA

Satellite imagery is an important source of raster data, with multiple applications in science and economy. Historical imagery together with growing archives of active satellite missions present big amounts of data. Sentinel-2 mission launched in 2015. is the last in row of satellite missions aimed at collecting global high resolution images with high time frequency. Data from this mission is freely available for scientific and commercial uses. Handling and analysing that data efficiently presents a challenge due to the size of the data – according to Sentinel-2 annual report for 2017, 10TB of data is generated daily, with size of overall data collected from this mission since launch exceeding 40PB. Raster databases offer a solution for fast, efficient and reliable handling of big amounts of raster data. This paper describes the results of testing rasdaman raster database, using a timeseries of Sentinel-2 images over test area.

Key words: time series, Sentinel-2, raster databases