Uvod u vrste i formate podataka o okolišu

Ishodi učenja

Nakon završetka ove nastavne cjeline studenti će moći:

- definirati osnovne vrste podataka o okolišu i objasniti njihovu važnost u istraživanjima okoliša;
- razlikovati strukturirane, polustrukturirane i nestrukturirane podatke o okolišu;
- prepoznati uobičajene formate podataka u znanostima o okolišu (CSV, NetCDF, HDF5, GeoJSON, Shapefile i dr.);
- procijeniti prednosti i ograničenja različitih formata pri pohrani, razmjeni i analizi podataka;
- pretvarati datoteke s podacima o okolišu iz jednog formata u drugi i obrađivati ih odgovarajućim alatima i bibliotekama (npr. Python i QGIS).

Uvod

Podaci o okolišu imaju ključnu ulogu u razumijevanju, praćenju i upravljanju prirodnim sustavima te sustavima na koje utječe čovjek. Oni pružaju empirijsku osnovu za proučavanje klimatskih promjena, kvalitete zraka i vode, bioraznolikosti, načina korištenja zemljišta i brojnih drugih procesa u okolišu. Napredak senzorskih tehnologija, satelitskih sustava za promatranje Zemlje i računalnih alata doveo je do dosad nezabilježenog povećanja količine, raznolikosti i složenosti podataka o okolišu. Stoga stručnjaci i studenti u području znanosti o okolišu trebaju znati analizirati podatke te razumjeti njihove vrste i formate u kojima se pohranjuju i razmjenjuju.

Podaci o okolišu prikupljaju se iz brojnih izvora. Mjerenja na terenu provode se zemaljskim mjernim instrumentima, kao što su meteorološke postaje, senzori za praćenje kvalitete zraka, vodomjerne postaje i plohe za ekološko praćenje. Platforme za daljinska istraživanja, uključujući satelite, bespilotne letjelice i zrakoplove, omogućuju prostorno obuhvatna opažanja Zemljine površine i atmosfere. Podaci mogu nastati i primjenom numeričkih modela i simulacija, primjerice klimatskih ili hidroloških modela, koji stvaraju velike višedimenzionalne skupove podataka. Svaki od tih izvora daje podatke različite strukture, mjerila i tehničkih zahtjeva, što utječe na njihovu pohranu i uporabu.

Razumijevanje vrsta i formata podataka o okolišu nužno je za učinkovito rukovanje podacima i njihovu analizu. Vrste podataka opisuju način organizacije informacija, primjerice u obliku strukturiranih tabličnih podataka, polustrukturiranih zapisa ili nestrukturiranih datoteka poput slika

i tekstova. Formati podataka određuju način kodiranja podataka u datotekama kao što su CSV, Parquet, NetCDF, HDF5, GeoTIFF i GeoJSON. Odabir formata utječe na dostupnost podataka, učinkovitost pohrane, interoperabilnost programskih alata i mogućnost uključivanja metapodataka o podrijetlu podataka, mjernim jedinicama, prostornom referentnom sustavu i vremenskom obuhvatu.

Nepravilno rukovanje formatima podataka u znanostima o okolišu može dovesti do pogrešnog tumačenja, gubitka podataka ili slabije ponovljivosti rezultata. Primjerice, zanemarivanje koordinatnih referentnih sustava prostornih podataka može uzrokovati neispravno preklapanje kartografskih slojeva, dok uporaba prejednostavnih formata može rezultirati gubitkom važnih metapodataka. Nasuprot tomu, odgovarajući formati omogućuju učinkovitu razmjenu, dugoročno očuvanje i objedinjavanje raznorodnih skupova podataka iz više izvora. To je osobito važno u interdisciplinarnim istraživanjima okoliša, u kojima je potrebno povezati i usporediti podatke iz različitih područja.

U ovom se predavanju predstavljaju glavne vrste podataka o okolišu i formati koji se najčešće primjenjuju u znanostima o okolišu i geoprostornim aplikacijama. Objašnjava se kako su podaci organizirani i pohranjeni, uz praktične primjere praćenja kvalitete zraka, klimatskih opažanja i analize korištenja zemljišta. Usvajanjem tih koncepata studenti će biti spremniji za rad sa stvarnim skupovima podataka, odabir prikladnih formata za analizu te uspostavu transparentnih, ponovljivih i znanstveno utemeljenih postupaka obrade podataka.

Vrste podataka o okolišu

Podaci o okolišu mogu se klasificirati prema strukturi, dimenzionalnosti, prostornim i vremenskim obilježjima te načinu prikupljanja. Poznavanje njihovih vrsta važno je za odabir odgovarajućih formata pohrane, analitičkih postupaka i načina vizualizacije. U ovom se poglavlju podaci prema strukturnim obilježjima dijele u tri osnovne kategorije: strukturirane, polustrukturirane i nestrukturirane podatke. Posebna se pozornost pridaje prostornoj i vremenskoj dimenziji, koje su iznimno važne u primjenama povezanima s okolišem.

Strukturirani podaci o okolišu

Strukturirani podaci jasno su organizirani i najčešće pohranjeni u tablicama, u kojima su raspoređeni u retke i stupce s precizno određenim atributima i tipovima podataka. Svaki redak obično predstavlja jedno opažanje, a svaki stupac jednu varijablu ili atribut. Takvi su podaci česti

u programima praćenja stanja okoliša, a najčešće se prikupljaju standardiziranim mjernim postupcima.

Primjeri strukturiranih podataka o okolišu uključuju vremenske nizove temperature zraka, mjerenja oborina, koncentracije onečišćujućih tvari u zraku, pokazatelje kvalitete vode i svojstva tla izmjerena na stalnim lokacijama. Ti se skupovi podataka često pohranjuju u CSV datotekama, proračunskim tablicama ili relacijskim bazama podataka. Strukturirane je podatke razmjerno jednostavno pretraživati, filtrirati i analizirati statističkim programima, sustavima baza podataka te programskim jezicima kao što su R i Python.

CSV je jednostavan tekstni format u kojem se podaci zapisuju po recima. Lako se čita i razmjenjuje, ali nije osobito učinkovit za opsežnu analitičku obradu ni dugoročnu pohranu velikih količina podataka. Nasuprot tomu, Apache Parquet stupčano je organiziran format optimiziran za analitičke zadatke. Omogućuje brže upite nad odabranim varijablama, učinkovito sažimanje i očuvanje tipova podataka. Stoga je Parquet prikladniji za velike skupove podataka o okolišu i podatkovno zahtjevne postupke obrade, dok je CSV i dalje koristan za manje skupove i jednostavnu razmjenu podataka.

Glavne su prednosti strukturiranih podataka jednostavnost i kompatibilnost s brojnim alatima. Omogućuju učinkovito objedinjavanje podataka, statističku analizu i vizualizaciju, primjerice izračun srednjih vrijednosti, trendova i korelacija tijekom vremena. Ipak, imaju i ograničenja: manje su prikladni za prikaz složenih prostornih struktura, višedimenzionalnih varijabli ili velikih rasterskih skupova podataka. Osim toga, važan kontekst može se izgubiti ako metapodaci nisu pažljivo dokumentirani i zasebno pohranjeni.

Polustrukturirani podaci o okolišu

Polustrukturirani podaci ne slijede strogo zadanu tabličnu strukturu, ali sadržavaju organizacijske elemente poput oznaka, ključeva ili hijerarhijskih odnosa. Fleksibilniji su od strukturiranih podataka, a pritom zadržavaju dovoljnu razinu organizacije za automatiziranu obradu.

U znanostima o okolišu polustrukturirani podaci često služe za prikaz metapodataka, izlaznih podataka senzora i prostornih objekata. Formati XML i JSON široko se primjenjuju za razmjenu podataka među sustavima, osobito u web-uslugama i aplikacijskim programskim sučeljima (API). GeoJSON, specijalizirani format temeljen na JSON-u, često se koristi za prikaz geografskih objekata zajedno s njihovim atributima i geometrijom.

Polustrukturirani podaci osobito su korisni pri radu s raznorodnim skupovima podataka različite strukture ili kada se tijekom vremena mogu dodavati novi atributi. Primjerice, mreže senzora za praćenje okoliša često stvaraju polustrukturirane zapise koji uključuju izmjerene vrijednosti, vremenske oznake, identifikatore senzora i oznake kvalitete. Ta fleksibilnost olakšava povezivanje različitih platformi, ali u usporedbi sa strukturiranim tabličnim podacima može zahtijevati složenije raščlanjivanje i provjeru valjanosti.

Nestrukturirani podaci o okolišu

Nestrukturirani podaci nemaju unaprijed definiran podatkovni model i nije ih jednostavno prikazati u tablicama ili hijerarhijskim strukturama. U ovu kategoriju ubrajaju se slike, videozapisi, zvučne snimke, izvješća u slobodnom tekstnom obliku i izvorne snimke prikupljene daljinskim istraživanjima. U području okoliša njihova važnost raste zahvaljujući napretku tehnologija promatranja Zemlje i sustava za automatizirano prikupljanje podataka.

Primjeri nestrukturiranih podataka uključuju satelitske snimke, zračne fotografije, snimke bespilotnih letjelica, skenirane terenske bilješke i tekstualne studije procjene utjecaja na okoliš. Za njihovu su obradu često potrebne specijalizirane metode, kao što su analiza slike, obrada prirodnog jezika i strojno učenje. Iako mogu sadržavati vrlo bogate i detaljne informacije, zahtjevniji su za analizu i učinkovitu pohranu.

Analiza nestrukturiranih podataka o okolišu često zahtijeva dodatno tumačenje ili pretvorbu. Primjerice, obradom satelitskih snimaka mogu se izračunati vegetacijski indeksi ili razredi pokrova zemljišta, čime se izvorni rasterski podaci pretvaraju u strukturirane ili polustrukturirane skupove prikladne za daljnju analizu. Nestrukturirani podaci zato često služe kao polazni materijal za stvaranje organiziranijih informacija o okolišu.

Prostorni podaci o okolišu

Prostorni podaci odnose se na informacije izravno povezane s geografskim položajem. Položaj se može odrediti koordinatama, administrativnim granicama, mrežama ili prostornim referentnim sustavima. Prostorni su podaci temeljni za znanosti o okolišu jer je većina pojava u okolišu neraskidivo povezana s prostorom.

Prostorni podaci o okolišu najčešće se dijele na vektorske i rasterske. Vektorski podaci prikazuju diskretne objekte, poput mjernih postaja, rijeka, prometnica i zemljišnih čestica, pomoću točaka, linija i poligona. Rasterski podaci prikazuju kontinuirane pojave, poput temperaturnih polja, modela nadmorske visine ili satelitskih snimaka, mrežom ćelija odnosno piksela.

Razumijevanje prostornih podataka presudno je za analizu okoliša i izradu karata. Koordinatni referentni sustavi, prostorna razlučivost i mjerilo mogu znatno utjecati na rezultate analize. Nepravilno rukovanje prostornim podacima može dovesti do nepodudaranja slojeva, netočnih mjerenja ili pogrešnih zaključaka. Zato njihovo upravljanje zahtijeva posebnu pozornost i specijalizirane alate, kao što su geografski informacijski sustavi (GIS).

Vremenski podaci o okolišu

Vremenska obilježja još su jedna važna odrednica podataka o okolišu. Brojni se procesi mijenjaju tijekom vremena, od kratkotrajnih promjena vremenskih prilika do dugoročnih klimatskih trendova i promjena ekosustava. Podaci se mogu prikupljati u pravilnim ili nepravilnim vremenskim razmacima ili kontinuiranim mjerenjem.

Vremenski nizovi čest su oblik vremenskih podataka o okolišu. Primjeri su satna mjerenja kvalitete zraka, dnevne vrijednosti protoka rijeka i mjesečni klimatski indeksi. Pri radu s takvim podacima treba voditi računa o vremenskoj razlučivosti, načinu označavanja vremenskih zona i nedostajućim vrijednostima. Pravilno vremensko usklađivanje ključno je pri povezivanju skupova podataka iz različitih izvora.

Podaci o okolišu često imaju i prostornu i vremensku dimenziju pa nastaju prostorno-vremenski skupovi podataka. Oni mogu biti složeni i visoke dimenzionalnosti, primjerice izlazni podaci globalnih klimatskih modela ili vremenski nizovi dobiveni satelitskim opažanjima. Za njihovu obradu potrebne su specijalizirane strukture i formati koji omogućuju učinkovitu pohranu i pristup velikim količinama informacija.

Višedimenzionalni i mrežno organizirani podaci

Višedimenzionalni podaci opisuju varijable okoliša koje se mijenjaju duž više dimenzija, primjerice geografske širine, geografske dužine, nadmorske visine i vremena. Česti su u atmosferskim i oceanografskim istraživanjima, u kojima se mjerenja provode u trodimenzionalnom prostoru tijekom duljih vremenskih razdoblja.

Podaci na pravilnoj mreži posebna su vrsta višedimenzionalnih podataka čije su vrijednosti zadane u čvorovima ili ćelijama pravilne mreže. Primjeri su skupovi podataka klimatskih reanaliza, izlazni podaci numeričkih modela za prognozu vremena i modeli oceanske cirkulacije. Takvi skupovi često sadržavaju više varijabli i opsežne metapodatke pa su znatno složeniji od jednostavnih tabličnih podataka.

Složenost višedimenzionalnih podataka o okolišu zahtijeva specijalizirane formate i programske alate koji učinkovito obrađuju velike nizove i čuvaju metapodatke. Prije odabira formata i postupaka analize potrebno je razumjeti prirodu tih podataka.

Uobičajeni formati podataka o okolišu

Formati podataka o okolišu određuju način kodiranja, pohrane i razmjene podataka među sustavima. Odabir formata bitno utječe na dostupnost, učinkovitost pohrane, interoperabilnost i dugoročnu uporabljivost podataka. U znanostima o okolišu formati često moraju podržavati velike količine podataka, prostorne i vremenske dimenzije te detaljne metapodatke. U ovom se potpoglavlju predstavljaju najčešći formati u okolišnim i geoprostornim primjenama te njihove značajke, prednosti i ograničenja.

Tablični formati podataka

Tablični formati među najraširenijima su u znanostima o okolišu zahvaljujući jednostavnosti i širokoj podršci različitih alata. Podatke pohranjuju u retke i stupce, a najčešće se koriste za strukturirane skupove poput vremenskih nizova mjerenja i zbirnih statističkih pokazatelja.

Najčešći je tablični format CSV (Comma-Separated Values), odnosno vrijednosti odvojene zarezima. CSV datoteke obične su tekstne datoteke u kojima svaki redak predstavlja zapis, a vrijednosti su odvojene zarezima ili drugim razdjelnicima. Format je jednostavan za stvaranje, čitanje i razmjenu među različitim programima. Često služi za razmjenu mjerenja kvalitete zraka, meteoroloških opažanja i podataka o kvaliteti vode. Međutim, CSV nema standardiziranu podršku za metapodatke, tipove podataka i prostorne referentne informacije, što bez odgovarajuće dokumentacije može izazvati nejasnoće ili gubitak podataka.

Često se koriste i formati proračunskih tablica, primjerice datoteke programa Microsoft Excel, osobito u početnim fazama prikupljanja podataka ili pri radu s manjim skupovima. Iako nude pristupačno korisničko sučelje i osnovne mogućnosti provjere valjanosti podataka, manje su prikladni za velike skupove, automatiziranu obradu i ponovljive postupke. Skriveno oblikovanje, više radnih listova i ručne izmjene mogu uzrokovati pogreške i smanjiti transparentnost.

Stupčani i analitički formati podataka

Apache Parquet otvoreni je stupčani format pohrane podataka, optimiziran za analitičke zadatke i obradu velikih količina podataka. Za razliku od tradicionalnih tabličnih formata, kao što

je CSV, koji pohranjuju podatke po recima, Parquet ih pohranjuje po stupcima. Time znatno ubrzava upite koji pristupaju samo dijelu varijabli te je osobito prikladan za velike skupove podataka o okolišu s brojnim atributima i opažanjima.

U znanostima o okolišu Parquet se sve češće koristi za pohranu velikih količina tabličnih podataka, kao što su dugoročna mjerenja kvalitete zraka, izlazni podaci klimatskih modela pretvoreni u tablični oblik i objedinjeni podaci senzora. Podržava učinkovito sažimanje i kodiranje, čime smanjuje potreban prostor za pohranu uz brzo čitanje. Čuva i informacije o tipovima podataka i shemi, čime otklanja mnoge nedostatke običnih tekstnih formata.

Iako Parquet nije izvorni GIS format, ima važnu ulogu u postupcima obrade podataka o okolišu na platformama u oblaku, u skladištima podataka i distribuiranim računalnim okruženjima, prije vizualizacije ili prostorne analize u GIS programima. Podaci pohranjeni u formatu Parquet mogu se pretvoriti u prostorne formate ili povezati s GIS alatima preko međukoraka obrade u Pythonu, sustavima temeljenima na SQL-u ili bibliotekama za prostorne podatke.

Znanstveni i višedimenzionalni formati podataka

Znanstveni formati osmišljeni su za učinkovitu pohranu velikih višedimenzionalnih skupova podataka te bogatih metapodataka o strukturi, mjernim jedinicama i koordinatnim sustavima. Osobito su važni u klimatologiji, meteorologiji, oceanografiji i istraživanjima atmosfere.

NetCDF (Network Common Data Form) jedan je od najčešćih znanstvenih formata u istraživanjima okoliša. Podržava višedimenzionalne nizove i samoopisne metapodatke pa je prikladan za prostorno-vremenske podatke, poput izlaznih podataka klimatskih modela, klimatskih reanaliza i varijabli izvedenih iz satelitskih opažanja. Jedna NetCDF datoteka može sadržavati više varijabli i dimenzija, uz učinkovit pristup podacima i izdvajanje njihovih podskupova.

HDF5 (Hierarchical Data Format, inačica 5) još je jedan snažan format za pohranu složenih skupova podataka o okolišu. Ima hijerarhijsku strukturu koja omogućuje organizaciju podataka u grupe i skupove, slično datotečnom sustavu unutar jedne datoteke. HDF5 je skalabilan i učinkovit te prikladan za vrlo velike skupove podataka. No njegova složenost može otežati rad početnicima, a čitanje i zapisivanje HDF5 datoteka često zahtijeva specijalizirane biblioteke.

Formati vektorskih prostornih podataka

Vektorski formati prikazuju diskretne geografske objekte pomoću točaka, linija i poligona. Neophodni su za kartiranje lokacija mjernih postaja, administrativnih granica, rijeka, prometnica i obilježja korištenja zemljišta.

ESRI Shapefile već je godinama standardni vektorski format u GIS-u. Sastoji se od više povezanih datoteka koje zajedno pohranjuju geometriju, atribute i podatke o prostornom referentnom sustavu. Iako ga podržava velik broj alata, ima više ograničenja, među ostalim glede veličine datoteka, duljine naziva atributa i dostupnih tipova podataka.

GeoJSON suvremen je, otvoren tekstni format za zapis geografskih objekata temeljen na JSON-u. Zahvaljujući jednostavnosti i čitljivosti osobito je popularan u web-kartografiji i razmjeni podataka. Omogućuje pohranu atributa i geometrije u jednoj datoteci te se dobro povezuje s web-tehnologijama. Međutim, kod složenih ili opsežnih skupova GeoJSON datoteke mogu postati vrlo velike i neučinkovite.

Formati rasterskih prostornih podataka

Rasterski formati koriste se za prikaz kontinuiranih pojava u okolišu, pri čemu se vrijednosti pohranjuju u mreži ćelija ili piksela. Tipični su primjeri satelitske snimke, digitalni modeli reljefa, karte temperature Zemljine površine i vegetacijski indeksi.

GeoTIFF široko je rasprostranjen rasterski format koji nadograđuje standardni TIFF uključivanjem geografskih metapodataka, primjerice koordinatnog referentnog sustava i prostorne razlučivosti. Često se primjenjuje u daljinskim istraživanjima i GIS-u, a podržava ga većina geoprostornih programskih alata.

Za mrežno organizirane podatke o okolišu često se koriste i drugi formati, poput NetCDF-a i HDF5-a, osobito kada podaci uključuju vremensku ili visinsku dimenziju. Odabir rasterskog formata ovisi ponajprije o složenosti podataka, veličini datoteka i zahtjevima analize.

Metapodaci i standardi opisivanja podataka

Metapodaci su ključni za pravilno razumijevanje i korištenje podataka o okolišu. Opisuju njihovo podrijetlo, sadržaj, strukturu i kvalitetu te prostorni i vremenski obuhvat. Bez dostatnih metapodataka skupove podataka može biti teško, pa čak i nemoguće, ispravno protumačiti.

U okolišnim i geoprostornim primjenama koristi se više standarda metapodataka. ISO 19115 međunarodni je standard za geografske metapodatke koji omogućuje sveobuhvatno opisivanje

prostornih skupova podataka. Dublin Core jednostavniji je i općenitiji standard koji se primjenjuje u različitim područjima. Mnogi znanstveni formati, primjerice NetCDF, podržavaju ugrađene metapodatke usklađene s uspostavljenim konvencijama.

Pravilna primjena standarda metapodataka poboljšava interoperabilnost, ponovljivost rezultata i dugoročno očuvanje podataka. Ujedno olakšava pronalaženje podataka na portalima i u repozitorijima te omogućuje istraživačima i donositeljima odluka njihovu učinkovitu ponovnu uporabu.

Odabir odgovarajućeg formata podataka

Odabir prikladnog formata važan je korak u upravljanju podacima o okolišu. Ovisi o veličini i strukturi podataka, njihovim prostornim i vremenskim dimenzijama, ciljevima analize i korisnicima kojima su namijenjeni. Jednostavni tablični formati mogu biti dovoljni za male i dobro dokumentirane skupove, dok su za opsežno modeliranje i analizu okoliša potrebni složeniji višedimenzionalni formati.

Postupci obrade podataka o okolišu često uključuju pretvorbu između formata radi ispunjavanja određenih zahtjeva analize ili vizualizacije. Poznavanje njihovih prednosti i ograničenja omogućuje informiran odabir te smanjuje rizik od gubitka ili pogrešnog tumačenja podataka.

Primjer primjene: istraživanje formata podataka o okolišu u QGIS-u

Geografski informacijski sustavi (GIS) pružaju napredno okruženje za rad s prostornim podacima o okolišu iz različitih izvora i u različitim formatima. QGIS je široko korištena GIS platforma otvorenoga koda koja podržava brojne formate podataka o okolišu i analitičke postupke. Ovaj primjer pokazuje kako se u QGIS-u mogu istraživati, vizualizirati i pretvarati vrste i formati podataka o okolišu, uz isticanje praktičnih izazova i preporučenih postupaka rukovanja podacima.

Pregled primjera primjene

Primjer se usredotočuje na objedinjavanje i analizu podataka o okolišu pohranjenih u različitim formatima unutar jednog GIS projekta. Studenti rade s tri reprezentativna skupa: mjerenjima kvalitete zraka u CSV formatu, podacima o pokrovu zemljišta u vektorskom formatu (Shapefile ili GeoJSON) te podacima o temperaturi izvedenima iz satelitskih opažanja u rasterskom formatu (GeoTIFF ili NetCDF). Ti se skupovi razlikuju po strukturi, dimenzionalnosti i sadržaju metapodataka pa dobro ilustriraju koncepte iz prethodnih poglavlja.

Cilj primjera nije napredna prostorna analiza, nego stjecanje praktičnih vještina učitavanja, pregleda i upravljanja podacima o okolišu u različitim formatima. Radom u QGIS-u studenti uče kako formati utječu na vizualizaciju, prostorno usklađivanje i mogućnost tumačenja podataka.

Učitavanje tabličnih podataka o okolišu (CSV)

Prvi korak obuhvaća učitavanje mjerenja kvalitete zraka pohranjenih u CSV datoteci. Skup podataka obično sadržava stupce s identifikatorom postaje, geografskom širinom, geografskom dužinom, datumom i koncentracijom onečišćujuće tvari. Pri uvozu CSV podataka u QGIS studenti trebaju odrediti stupce s prostornim koordinatama te odabrati odgovarajući koordinatni referentni sustav.

Ovaj korak pokazuje važne aspekte rada s tabličnim podacima o okolišu. Pogrešno definirane koordinate mogu dovesti do neispravnog položaja točaka, a nedostatak metapodataka može zahtijevati pretpostavke o mjernim jedinicama ili mjernim standardima. Nakon učitavanja CSV podaci prikazuju se kao točkasti objekti, što studentima omogućuje vizualni pregled prostornog rasporeda mjernih postaja.

Studenti trebaju istražiti atributne tablice, primijeniti jednostavne filtre i pregledati vremenske obrasce upitima nad atributima. Time se pokazuje kako se strukturiranim podacima mogu pridružiti prostorne informacije i uključiti ih u GIS postupke obrade.

Rad s vektorskim podacima o okolišu

Drugi skup sadržava informacije o pokrovu zemljišta u vektorskom formatu, primjerice Shapefile ili GeoJSON. Prikazuje diskretne prostorne objekte i sadržava atribute koji opisuju kategorije korištenja zemljišta. Nakon učitavanja u QGIS studenti pregledavaju vrste geometrije, atributna polja i koordinatni referentni sustav.

U ovom se dijelu vježbe naglašava važnost prostornih referentnih sustava i dosljednosti atributa. Studenti uspoređuju koordinatni referentni sustav podataka o pokrovu zemljišta sa sustavom točaka za mjerenje kvalitete zraka te po potrebi reprojiciraju slojeve. Istražuju i osnovne mogućnosti kartografskog prikaza kako bi vizualizirali razrede pokrova zemljišta i razumjeli povezanost vektorskih atributa s prostornim objektima.

Tako studenti upoznaju prednosti i ograničenja vektorskih formata, uključujući strukturu datoteka, pohranu atributa i interoperabilnost.

Istraživanje rasterskih podataka o okolišu

Treći skup podataka rasterski je prikaz površinske temperature dobivene satelitskim opažanjima ili neke druge kontinuirane varijable okoliša. Može biti pohranjen u formatu GeoTIFF ili NetCDF. Rasterski se podaci u QGIS učitavaju kao slojevi temeljeni na mreži, pri čemu svaka ćelija sadržava brojčanu vrijednost izmjerene ili modelirane varijable.

Studenti pregledavaju svojstva rastera, kao što su prostorna razlučivost, prostorni obuhvat i tip podataka. Prostorne obrasce tumače primjenom vizualizacijskih tehnika, uključujući gradijente boja i razvrstavanje vrijednosti. Ako se koristi NetCDF, istražuju kako unutar jedne datoteke pristupiti različitim varijablama ili vremenskim koracima.

Ovaj korak pokazuje zašto su rasterski formati prikladni za prikaz kontinuiranih pojava u okolišu te kako ugrađeni metapodaci omogućuju ispravno prostorno tumačenje.

Usporedba i integracija različitih formata podataka

Nakon učitavanja svih skupova podataka studenti preklapaju vektorske i rasterske slojeve u istom projektu QGIS-a. Tako mogu vizualno procijeniti odnose između mjerenja kvalitete zraka, vrsta pokrova zemljišta i temperaturnih obrazaca. Praktična integracija različitih vrsta podataka može se pokazati jednostavnim prostornim upitima, primjerice pronalaženjem mjernih postaja unutar određenih razreda pokrova zemljišta.

Usporedba pokazuje kako se formati podataka o okolišu međusobno nadopunjuju i kako GIS služi kao jedinstvena platforma za analizu podataka različitih formata. Studenti pritom uočavaju i utjecaj razlika u kvaliteti i razlučivosti podataka na njihovo tumačenje.

Pretvorba i izvoz podataka

Važan dio primjera jest pretvorba podataka. Studenti vježbaju izvoz vektorskih slojeva iz formata Shapefile u GeoJSON te izvoz rasterskih slojeva iz GeoTIFF-a u drugi podržani format. Time se učvršćuje spoznaja da odabir formata često ovisi o konkretnom zadatku, zahtjevima programa ili potrebama razmjene podataka.

Posebna se pozornost pridaje očuvanju metapodataka tijekom pretvorbe i razumijevanju mogućih ograničenja i gubitaka podataka. Studente se potiče na dokumentiranje koraka pretvorbe i provjeru izlaznih datoteka.

Praktične vježbe — zadaci u QGIS-u korak po korak

Ove praktične vježbe osmišljene su radi stjecanja iskustva u radu s vrstama i formatima podataka o okolišu pomoću QGIS-a, besplatnoga geografskog informacijskog sustava otvorenoga koda. Studenti će raditi sa stvarnim skupovima podataka o okolišu u različitim formatima te učiti kako ih učitati, vizualizirati, povezati i pretvoriti.

Izvori podataka (besplatan i otvoren pristup)

1. Podaci o praćenju kvalitete zraka (CSV)

Besplatni podaci o kvaliteti zraka i mjernim postajama dostupni su preko Europske agencije za okoliš (EEA) i drugih državnih portala otvorenih podataka.

• Europska agencija za okoliš — podaci i karte okoliša

https://www.eea.europa.eu/Data and Maps

Studenti mogu preuzeti mjerenja kvalitete zraka i izvesti ih u CSV datoteke koje sadržavaju koordinate postaja, vremenske oznake i koncentracije onečišćujućih tvari.

Alternativna mogućnost:

Dodatak NASA POWER za QGIS omogućuje izravan pristup klimatskim i atmosferskim varijablama (temperatura, oborine, zračenje), bez ručnog preuzimanja datoteka.

---

2. Podaci o pokrovu zemljišta (raster / GeoTIFF ili web-usluge)

ESA WorldCover pruža globalni skup podataka o pokrovu zemljišta izveden iz satelitskih snimaka misije Sentinel.

• ESA WorldCover (preuzimanje ili web-usluge)

http://ESA WorldCover.org

Usluge WMTS/WMS (mogu se izravno koristiti u QGIS-u):
• WMTS: https://services.terrascope.be/wmts/v2
• WMS: https://services.terrascope.be/wms/v2

---

3. Satelitske snimke (raster / GeoTIFF)

Besplatne satelitske snimke prikladne za analize okoliša mogu se preuzeti sa sljedećih izvora:

• Copernicus Data Space (misije Sentinel)
https://dataspace.copernicus.euhttps://dataspace.copernicus.eu

• USGS EarthExplorer (arhiva Landsat)
https://earthexplorer.usgs.govhttps://earthexplorer.usgs.gov

Zbog veličine datoteka preporučuje se preuzimanje manjih prostornih podskupova ili pločica (engl. tiles).

---

4. Vektorski podaci (OpenStreetMap)

OpenStreetMap pruža besplatne vektorske podatke o prometnicama, korištenju zemljišta, granicama i infrastrukturi.

• OpenStreetMap

https://www.openstreetmap.org

Podaci se u QGIS mogu uvesti pomoću dodatka QuickOSM.

Vježba 1 — Uvoz i vizualizacija CSV podataka o kvaliteti zraka

Cilj: Učitati tablične podatke o kvaliteti zraka i prikazati mjerne postaje kao točkaste objekte.

Postupak:

1. Otvorite QGIS i odaberite Layer → Add Layer → Add Delimited Text Layer.
2. Pronađite preuzetu CSV datoteku (npr. air_quality.csv).
3. Za polje X odaberite geografsku dužinu, a za polje Y geografsku širinu.
4. Odaberite koordinatni referentni sustav (EPSG:4326).
5. Kliknite Add kako biste prikazali točke na karti.
6. Otvorite Attribute Table (F6) i pregledajte dostupne varijable.

Rasprava:
Koji su atributi uključeni u skup podataka? Jesu li mjerne jedinice i vremenske oznake jasno definirane?

Vježba 2 — Učitavanje i oblikovanje prikaza podataka

Cilj: Vizualizirati podatke o pokrovu zemljišta pomoću rasterskih slojeva i primijeniti tematski prikaz.

Postupak:

1. U QGIS-u odaberite Layer → Add Layer → Add WMS/WMTS Layer.
2. Uspostavite novu vezu koristeći WMTS URL usluge ESA WorldCover.
3. Odaberite sloj pokrova zemljišta i dodajte ga projektu.
4. Otvorite Layer Styling (F7) i primijenite kategorijsku shemu boja za razrede pokrova zemljišta.
5. Dodajte legendu za tumačenje kategorija pokrova zemljišta.

Rasprava:
Kako se prostorni raspored pokrova zemljišta razlikuje unutar promatranog područja?

Vježba 3 — Učitavanje i istraživanje satelitskih rasterskih podataka

Cilj: Učitati satelitske rasterske podatke i istražiti prostorne obrasce.

Postupak:

1. Preuzmite snimku Sentinel-2 ili Landsat u formatu GeoTIFF s portala Copernicus ili USGS.
2. U QGIS-u odaberite Layer → Add Layer → Add Raster Layer i učitajte datoteku.
3. Pregledajte svojstva rastera, kao što su razlučivost, obuhvat i koordinatni sustav.
4. Primijenite gradijent boja ili izradite kompozit u lažnim bojama.
5. (Neobvezno) Koristite Raster Calculator za izračun normaliziranog diferencijskog vegetacijskog indeksa (NDVI).

Rasprava:
Koja se obilježja okoliša mogu uočiti u rasterskim podacima?

Vježba 4 — Integracija vektorskih i rasterskih slojeva

Cilj: Povezati više formata podataka i provesti osnovno prostorno tumačenje.

Postupak:

1. Preklopite točke mjerenja kvalitete zraka sa slojevima pokrova zemljišta i satelitskih rastera.
2. Alatom Identify Features pregledajte vrijednosti rastera na lokacijama točaka.
3. (Neobvezno) Koristite Raster → Zonal Statistics za sažimanje rasterskih vrijednosti po razredima pokrova zemljišta.

Rasprava:
Mogu li se uočiti povezanosti između izmjerene kvalitete zraka i vrsta pokrova zemljišta?

Vježba 5 — Pretvorba i izvoz formata podataka

Cilj: Pretvarati podatke o okolišu iz jednog formata u drugi.

Postupak:

1. Desnom tipkom miša kliknite vektorski sloj i odaberite Export → Save Features As….
2. Izvezite sloj u format GeoJSON.
3. Za rasterske podatke odaberite Raster → Conversion → Translate (Convert Format).
4. Spremite raster u drugi format (npr. GeoTIFF ili GeoPackage).
5. Provjerite jesu li očuvani prostorni referentni sustav i atributi.

Rasprava:
Koji su formati najprikladniji za razmjenu i dugoročnu pohranu podataka?

Provjera znanja

Pitanje 1. Koja je vrsta podataka o okolišu najprikladnija za tablični prikaz u retcima i stupcima?

A) Nestrukturirani podaci

B) Polustrukturirani podaci

C) Strukturirani podaci

D) Rasterski podaci

Točan odgovor: C) Strukturirani podaci

Pitanje 2. Koji je format najprikladniji za pohranu višedimenzionalnih klimatskih ili atmosferskih podataka s ugrađenim metapodacima?

A) CSV
B) GeoJSON
C) NetCDF
D) Shapefile

Točan odgovor: C) NetCDF

Pitanje 3. Koja se vrsta podataka u QGIS-u prvenstveno koristi za prikaz kontinuiranih pojava u okolišu, poput temperature ili nadmorske visine?

A) Vektorski podaci (točke, linije, poligoni)
B) Rasterski podaci

C) Tablični podaci

D) Tekstni podaci

Točan odgovor: B) Rasterski podaci

Pitanje 4. Koja je ključna prednost formata GeoJSON u odnosu na ESRI Shapefile?

A) Podržava veće datoteke od svih drugih formata

B) Pohranjuje geometriju i atribute u jednu, čovjeku čitljivu datoteku

C) Brži je za rastersku analizu

D) Obvezan je u svim GIS aplikacijama

Točan odgovor: B) Pohranjuje geometriju i atribute u jednu, čovjeku čitljivu datoteku

---

Pitanje 5. Zašto su metapodaci osobito važni u skupovima podataka o okolišu?

A) Smanjuju veličinu datoteke

B) Povećavaju brzinu vizualizacije

C) Opisuju podrijetlo podataka, mjerne jedinice te prostorne i vremenske reference

D) Zamjenjuju potrebu za analizom podataka

Točan odgovor: C) Opisuju podrijetlo podataka, mjerne jedinice te prostorne i vremenske reference

---

Praktični zadaci (4–5 vježbi)

Studente se vrednuje na temelju uspješnog izvršavanja sljedećih praktičnih vježbi u QGIS-u, uz uporabu otvorenih skupova podataka o okolišu.

1. Uvoz i pregled podataka
   Učitajte CSV skup mjerenja okoliša u QGIS, odredite ispravan koordinatni referentni sustav te pregledajte atributna polja i metapodatke.
2. Rad s vektorskim i rasterskim podacima
   Učitajte vektorske podatke o pokrovu zemljišta i rasterske satelitske snimke, provjerite njihovu prostornu usklađenost te primijenite odgovarajući kartografski prikaz.
3. Integracija podataka
   Preklopite točkaste, vektorske i rasterske skupove podataka te protumačite prostorne odnose među varijablama okoliša.
4. Pretvorba formata podataka
   Pretvorite vektorske podatke iz formata Shapefile u GeoJSON i izvezite rasterske podatke u drugi podržani format, uz očuvanje prostornih referentnih informacija.
5. Vizualizacija i tumačenje (neobvezno / dodatni zadatak)
   Izradite jednostavan kartografski prikaz s legendom i naslovom te ukratko protumačite uočene obrasce u okolišu.

Literatura i preporučeni izvori za daljnje čitanje

- Davis, Jerry. Introduction to Environmental Data Science. Bookdown.org, 2022. Otvoreni pristup. https://bookdown.org/igisc/EnvDataSci.https://bookdown.org/igisc/EnvDataSci
- Earth-Env-Data-Science Project. Earth & Environmental Data Science – An Introduction. 2023. Otvoreni obrazovni izvor. https://earth-env-data-science.github.io/intro.html.https://earth-env-data-science.github.io/intro.html
- Lovelace, Robin, Jakub Nowosad i Jannes Muenchow. Geocomputation with R. CRC Press, 2019. Izdanje u otvorenom pristupu. https://r.geocompx.org/.https://r.geocompx.org/
- Rey, Sergio J. i Levi J. Wolf. Geo-data science with Python. Geographicdata.science, 2022. Udžbenik u otvorenom pristupu. https://geographicdata.science/book/intro.html.https://geographicdata.science/book/intro.html
- Bolstad, Paul i Michael Shin. Essentials of Geographic Information Systems. 2. izd., University of Minnesota Libraries Publishing, 2018. Open Textbook Library. https://open.umn.edu/opentextbooks/textbooks/67.https://open.umn.edu/opentextbooks/textbooks/67