

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Teo Cerovski

**Avtomatsko ocenjevanje kakovosti
prostorskih podatkov**

DIPLOMSKO DELO
NA VISOKOŠOLSKEM STROKOVNEM ŠTUDIJU

Mentor: doc. dr. Rok Rupnik

Ljubljana, 2010



Št. naloge: 00460/2009

Datum: 01.09.2009

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **TEO CEROVSKI**

Naslov: **AVTOMATSKO OCENJEVANJE KAKOVOSTI PROSTORSKIH
PODATKOV**

AUTOMATED EVALUATION OF SPATIAL DATA QUALITY

Vrsta naloge: Diplomsko delo visokošolskega strokovnega študija

Tematika naloge:

Razložite pomen kakovosti prostorskih podatkov in naredite pregled pomembnih standardov s področja kakovosti prostorskih podatkov.

Raziščite katere kriterije kakovosti je možno oceniti avtomatsko in predstavite nekaj metod in algoritmov za takšno ocenjevanje. Na koncu z enim od orodij ocenite kakovost ene večjih podatkovnih zbirk prostorskih podatkov za območje Slovenije.

Mentor:

doc. dr. Rok Rupnik



Dekan:

prof. dr. Franc Solina

Rezultati diplomskega dela so intelektualna lastnina Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje Fakultete za računalništvo in informatiko ter mentorja.

Besedilo je oblikovano z urejevalnikom besedil L^AT_EX.

IZJAVA O AVTORSTVU

diplomskega dela

Spodaj podpisani Teo Cerovski,

z vpisno številko 63030143,

sem avtor diplomskega dela z naslovom:

Avtomatsko ocenjevanje kakovosti prostorskih podatkov

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom doc. dr. Roka Rupnika
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki "Dela FRI".

V Ljubljani, dne 15.03.2010

Podpis avtorja:

Zahvala

Zahvaljujem se mentorju, doc. dr. Roku Rupniku, za vse nasvete in vzpodbudo pri izdelavi tega diplomskega dela. Hvala Evi za skrbno lektoriranje, potrpežljivost in ker mi vedno stoji ob strani. Hvala tudi mami in očetu za vso podporo in spodbudo za dokončanje študija.

Kazalo

Povzetek	1
Abstract	2
1 Uvod	3
2 Prostorski podatki	5
2.1 Rastrski podatki	5
2.2 Vektorski podatki	5
3 Prostorski podatki in pomen kakovosti	6
4 Definicija koncepta kakovosti	8
4.1 Interna kakovost	9
4.2 Eksterna kakovost	12
5 Metapodatki o kakovosti prostorskih podatkov	13
6 Standardi na področju kakovosti prostorskih podatkov	14
6.1 ISO 19113:2002	14
6.2 ISO 19114:2003	15
6.3 ISO 19115:2003	15
6.4 ISO 19139:2007	16
6.5 INSPIRE	16
7 Metode za avtomatsko ocenjevanje interne kakovosti	17
7.1 Formatna usklajenost	18
7.2 Domenska in časovna usklajenost	19
7.3 Presežki in izostanki ter tematska natančnost	19
7.4 Pomenska usklajenost ali integriteta prostorskih podatkov	20
7.4.1 Prostorski relacijski operatorji	21
7.4.2 Kardinalnost	29
7.4.3 Pas evklidske oddaljenosti	30
7.5 Topološka usklajenost ali pravilnost prostorskih podatkov	32
7.5.1 Pogoste topološke napake in anomalije	33
7.5.2 Algoritmi za preverjanje topološke pravilnosti	38
7.6 Absolutna položajna natančnost	46
8 Ocenjevanje kakovosti digitalnega katastrskega načrta Slovenije	49

9	Zaključek	52
A	Slike topoloških napak in anomalij najdenih pri ocenjevanju DKN	53
	Seznam slik	59
	Seznam tabel	59
	Literatura	61

Seznam uporabljenih kratic

CEN	European Committee for Standardization (<i>Evropski odbor za standardizacijo</i>)
cSVT	client Spatial Validation Tool (<i>orodje za avtomatsko ocenjevanje kakovosti prostorskih podatkov</i>)
DE-9IM	Dimenzijsko razširjen 9-presečni model
DKN	Digitalni Katastrski Načrt
ESDIN	European Spatial Data Infrastructure best practice Network
FGDC	Federal Geographic Data Committee (<i>državni odbor ZDA za geografske podatke</i>)
GIS	Geografski Informacijski Sistem
GURS	Geodetska Uprava Republike Slovenije
INSPIRE	INfrastructure for SPatial InfoRmation in Europe (<i>Evropska infrastruktura za prostorske podatke</i>)
ISO	International Organization for Standardization (<i>Mednarodna organizacija za standardizacijo</i>)
JRC	Joint Research Center
KO	Katastrska Občina
OGC	Open Geospatial Consortium
RGB	Red, Green, Blue (<i>rdeča, zelena, modra</i>)
RMSE	Root Mean Square Error (<i>koren srednje kvadratne napake</i>)
SDI	Spatial Data Infrastructure (<i>infrastruktura za prostorske podatke</i>)

SIST	Slovenski Inštitut za Standardizacijo
SQL	Structured Query Language (<i>strukturiran povpraševalni jezik</i>)
TQM	Total Quality Management (<i>popolni nadzor nad kakovostjo</i>)
UML	Unified Modeling Language (<i>poenoten jezik modeliranja</i>)
XML	eXtensible Markup Language (<i>razširljiv označevalni jezik</i>)

Povzetek

Kakovost prostorskih podatkov je na področju geografske informatike vedno predstavljala precejšen problem. Z razmahom prostorskih informacijskih storitev na svetovnem spletu in vse večjo dostopnostjo prostorskih podatkov ter z uporabo le-teh za različne namene na mnogih področjih pa ta problem postaja vse bolj pereč.

Za objektivno ocenjevanje kakovosti prostorskih podatkov morajo obstajati jasno opredeljene metode oziroma postopki, ki jih je potrebno dosledno uporabljati. Standardi za kakovost prostorskih podatkov predpisujejo le splošne smernice in metodologije za ocenjevanje kakovosti, standardiziranje konkretnih metod pa je še v fazi preučevanja.

V diplomski nalogi je najprej predstavljena pomembnost kakovosti prostorskih podatkov in pomembnost ozaveščanja uporabnikov o tej kakovosti. V nadaljevanju je razložen uveljavljen koncept interne in eksterne kakovosti ter kriteriji za opisovanje kakovosti prostorskih podatkov. Na kratko je predstavljenih nekaj pomembnejših standardov iz obravnavanega področja. V osrednjem delu naloge je opisanih nekaj metod za avtomatsko ocenjevanje posameznih kriterijev kakovosti, s posebnim poudarkom na metodah in algoritmih za ocenjevanje pomenske in topološke usklajenosti. V zadnjem delu je prikazan primer avtomatskega ocenjevanja kakovosti na digitalnem katastrskem načrtu Slovenije.

Ključne besede:

prostorski podatki, interna kakovost, eksterna kakovost, avtomatsko ocenjevanje

Abstract

The quality of spatial data has always represented a significant problem in geomatics. This problem is becoming ever more pressing with the expansion of spatial information services on the Internet and with the increasing availability of spatial data and their use for various applications on many fields.

To objectively evaluate the quality of spatial data, clearly defined methods or procedures should exist and be applied consistently. Spatial data quality standards provide only general guidelines and methodologies for evaluating quality, while the standardization of specific methods is still being studied.

The thesis starts with presenting the importance of spatial data quality and the importance of user awareness on spatial data quality. Afterwards the concept of internal and external quality and the criteria for describing spatial data quality are explained. Some of the important standards in the field are presented briefly. The main section describes several methods for automated evaluation of individual quality criteria, with emphasis on methods and algorithms for evaluating the conceptual and topological consistency. The final section provides an example of automated quality evaluation on the digital cadastral plan of Slovenia.

Key words:

spatial data, internal quality, external quality, automated evaluation

1 Uvod

Tretjega februarja 1998 je ameriško vojaško letalo zadelo in pretrgalo nosilni kabel gondole v Cavaleaseju v Italiji in povzročilo 20 smrtnih žrtev. Letalo je imelo poškodovano krilo in rep, a se je uspelo vrniti v letalsko bazo. Preiskava je pokazala, da gondola ni bila vrisana na zemljevidu, ki ga je uporabljal pilot, merilnik višine letala pa je bil v okvari. Napake, pomanjkljivosti in negotovosti v prostorskih podatkih običajno ne vodijo do tako odmevnih mednarodnih incidentov, vendar je neizogibno dejstvo, da je kakršnokoli znanje o Zemljini površini podvrženo določenim negotovostim, pa naj bodo to lega pojavov v prostoru, njihov obstoj ali njihov opis. Nemogoče je poznati točno lokacijo česar koli na Zemljini površini, saj so merilne metode vedno podvržene mer-skim napakam, negotovosti se prikradejo na zemljevide, podatkovne baze in druge zapise skozi širok nabor postopkov pridobivanja in obdelave podatkov. Z ozirom na dejstvo, da dve osebi, ki kartirata isto območje, nikoli ne bosta izdelali identičnih kart, je večina klasifikacijskih shem za opis zemeljske površine, od sestave tal do rabe zemlje, samih po sebi negotovih.

Da bi uporabniki prostorskih podatkov lahko ocenili, do kakšne mere so določeni prostorski podatki primerni za njihovo nameravano uporabo, potrebujejo informacijo o kakovostnih značilnostih podatkov. Za primerjavo teh značilnosti med različnimi podatkovnimi nizi je najprej potreben standardiziran opis kakovosti in standardizirani kriteriji kakovosti. Ocenjevanje kriterijev kakovosti mora biti izvedeno z uporabo jasno opredeljenih metod in postopkov, ki so zaradi velike količine podatkov običajno v čim večji meri avtomatizirani. Končni rezultati ocenjevanja morajo biti podani v obliki standardnega poročila o kakovosti, običajno v obliki metapodatkov.

Žal standardi s področja ocenjevanja kakovosti predpisujejo le splošne smernice in metodologije za ocenjevanje kakovosti. Implementacija in uporaba konkretnih metod ocenjevanja je običajno stvar posameznih organizacij, kar pogosto povzroča težave pri primerjavi rezultatov ocenjevanja različnih proizvajalcev podatkov. Cilj te diplomske naloge je predstavitev nekaj teh metod za ocenjevanje posameznih kriterijev kakovosti.

V nadaljevanju je v drugem poglavju najprej podana kratka definicija prostorskih podatkov in delitev le-teh na rastrske in vektorske podatke. V tretjem poglavju je poudarjen pomen kakovosti prostorskih podatkov. V četrtem poglavju je definiran koncept kakovosti in delitev kakovosti na interno ter ek-

sterno. Našteti so elementi oziroma kriteriji za opisovanje kakovosti prostorskih podatkov. V petem poglavju so predstavljeni metapodatki kot način za opisovanje rezultatov ocenjevanja kakovosti. Šesto poglavje nudi pregled pomembnih standardov s področja kakovosti prostorskih podatkov in izpostavi nekaj pomanjkljivosti teh standardov. V sedmem, osrednjem poglavju je najprej razložen namen metod za ocenjevanje interne kakovosti prostorskih podatkov. Predstavljena je delitev metod ocenjevanja na neposredne in posredne ter na notranje in zunanje. Sledi opis metod za avtomatsko ocenjevanje posameznih kriterijev kakovosti. Posebej so poudarjene metode za ocenjevanje pomenske usklajenosti in algoritmi za ocenjevanje topološke usklajenosti, saj so te metode v primerjavi z ostalimi manj enostavne in tako bolj zanimive. V osmem poglavju je prikazan primer avtomatskega ocenjevanja kakovosti na digitalnem katastrskem načrtu (DKN) Slovenije. Ocenjevanje je izvedeno z orodjem cSVT, ki implementira v sedmem poglavju opisane metode ocenjevanja. V zaključku so povzete ugotovitve tega dela. Pomembnost kakovosti prostorskih podatkov je ponovno poudarjena.

2 Prostorski podatki

Prostorski podatki so “model realnosti” [2], razumska in poenostavljena predstavitev kompleksne realnosti, ki skušajo prikazati podobo prostorskega sveta, v kontekstu tega dela s tem mislimo predvsem geografski svet, svet, definiran s človeškimi izkustvi ter površino in bližnjo površino Zemlje.

Pojave (ali objekte) prostorskega sveta lahko razdelimo v abstrakciji diskretnih pojavov (hiša, jezero) in zveznih pojavov (onesnaženost zraka, relief), ki jih običajno opisujemo v obliki vektorskih in rastrskih podatkov.

2.1 Rastrski podatki

Rastrski podatki ali rastri so abstrakcija zveznih ali diskretnih pojavov realnega sveta, kjer so prostorski podatki razdeljeni v diskretne enote, predstavljene v obliki dvorazsežnega polja (matrike) celic. Vsaka celica vsebuje vrednost določene prostorske lastnosti (barva objekta, spektralni odboj, višina, temperatura). Prostorska lokacija celic je implicitno podana z njihovim zaporedjem v matriki. Matrika je predstavljena s kartezičnim koordinatnim sistemom, kjer so vrstice matrike vzporedne x-osi in stolpci y-osi kartezične ravnine.

Ločljivost celic je podana kot širina in višina celice v zemeljskih merskih enotah (npr. metrih). Dimenzije celice določajo, kako grobo ali fino bodo vzorci ali pojavi prikazani.

Vsaka celica hrani le eno vrednost, ki jo lahko razširimo z uporabo rastrskih pasov za predstavitev RGB barv, barvnimi tabelami (preslikava vrednosti celice v RGB vrednost) ali poljubnimi preslikovalnimi tabelami, kjer vrednosti celic predstavljajo ključne zapise v nekem drugem podatkovnem nizu.

Primeri rastrskih podatkov so digitalni letalski ali satelitski posnetki, fotografije, optično zajeti načrti, digitalni modeli reliefa, topografski zemljevidi ipd.

2.2 Vektorski podatki

Vektorski podatki so abstrakcija diskretnih pojavov realnega sveta, kjer so prostorski podatki predstavljeni z geometrijskimi oblikami in njihovimi opisnimi podatki. Geometrijske oblike, geometrijski objekti ali kar geometrije so

sestavljene iz množice koordinat, ki lahko predstavljajo točke ali zaporedja povezanih točk. V osnovi jih delimo na:

- *0-dimenzijske geometrijske objekte* ali *točke*, s katerimi običajno predstavljamo lokacije pojavov, katerih oblika ni pomembna ali ko je njihova velikost zelo majhna v primerjavi z okolico. Vsaka točka je opisana z njenimi koordinatami.
- *1-dimenzijske geometrijske objekte* ali *krivulje*, s katerimi opisujemo pojave linearne oblike, kot so ceste, reke ali daljnovodi. Osnovna oblika krivulje je *linija*, ki je sestavljena iz končne množice zaporedno povezanih točk. Dve povezani točki tvorita segment linije. Če sta prva in zadnja točka linije identični, govorimo o zaprti liniji ali linearnem obroču.
- *2-dimenzijske geometrijske objekte* ali *površine*, s katerimi opisujemo pojave, ki pokrivajo določeno površino, kot so parcele, jezera ali administrativne enote. Običajno jih predstavimo v obliki *poligona*. Poligon je območje, omejeno z natanko enim zunanjim linearnim obročem in nič ali več notranjimi obroči, ki predstavljajo luknje.
- *zbirke geometrijskih objektov*. Ko posameznega pojava ne moremo opisati z eno samo geometrijsko obliko, moramo le-te združiti v zbirko geometrijskih objektov, ki je tako tudi sama geometrijski objekt. Oblike ali objekti v zbirki imajo tako skupni identifikator in skupne opisne podatke.

Kot pri rastrskih podatkih tudi pri vektorskih sama geometrijska oblika pojava ne daje neke uporabne informacije. Zato moramo vsakemu pojavu poleg prostorske informacije podati tudi pripadajoče opisne ali atributne podatke. Vse podatke lahko shranjujemo v isti datoteki (ali tabeli), lahko pa imamo ločeni datoteki (ali tabeli) za prostorske in opisne podatke, kjer zapise pojavov povezujemo z enotnim identifikatorjem.

3 Prostorski podatki in pomen kakovosti

Vsak zemljevid ali prostorska baza je model, izdelan za določen namen, v katerem so bili določeni elementi spoznani za nebitvene in tako poenostavljeni, grupirani ali odstranjeni z namenom jasnejše predstavitve. Prostorski podatki so torej le približek realnega sveta in v določenih pogledih tako vsi neprecizni, nenatančni, nepopolni in zastareli.

z iskanjem ustrežnejših podatkov. Lahko pa sprejmejo negotovosti v podatkih in s tem potencialno tveganje in negativne posledice.

V zadnjih dveh desetletjih sta se zgodili dve pomembni revoluciji, ki sta močno povečali zanimanje za kakovost prostorskih podatkov. Prva je prehod iz papirnatih na digitalne zemljevide. Medtem ko so se uporabniki papirnatih zemljevidov zavedali težav z njihovo kakovostjo, jim digitalni format daje videz velike natančnosti, saj jim omogoča merjenje razdalj natančno na nekaj decimalnih mest in prikaz podatkov pri zelo visokem nivoju podrobnosti. Druga taka revolucija pa je razvoj Interneta in drugih komunikacijskih tehnologij, ki olajšujejo izmenjavo podatkov med posamezniki in organizacijami. Uporabniki si tako pogosto prenesejo prostorske podatke in jih uporabijo v lastne namene, brez da bi imeli kakršnekoli informacije o njihovi kakovosti.

Pojem “kakovost prostorskih podatkov” se običajno povezuje le s pozicijsko natančnostjo prostorskih podatkov. Vendar ima koncept kakovosti precej širši pomen in zajema celoten proces zajema, obdelave, hranitve, izmenjave in predstavitve prostorskih podatkov. Koncept kakovosti in elementi kakovosti so razdelani v naslednjem poglavju.

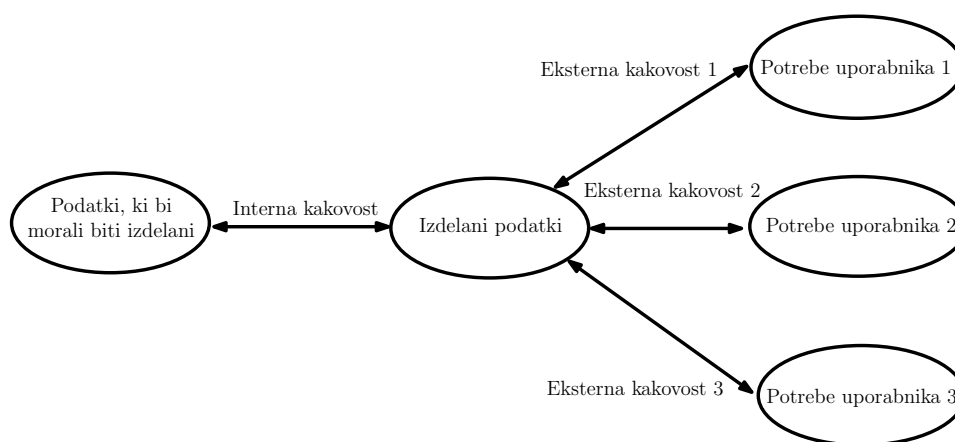
4 Definicija koncepta kakovosti

Od 20. stoletja naprej se pojem “kakovost” uporablja zlasti na področju proizvodnje za potrditev odličnosti izdelka. Prva dela na področju kakovosti se pripisujejo inženirju F.W. Taylorju, ki je opisal principe za upravljanje delovnega procesa z namenom izboljšanja kakovosti proizvodov. Prvi je te principe v svojih tovarnah uspešno vpeljal Ford. Po drugi svetovni vojni se je razvil koncept popolnega nadzora nad kakovostjo (*Total Quality Management* - TQM), ki so ga prvi vpeljali Japonci in s tem vzpodbudili industrijsko rast v državi [2]. Leta 1979 je organizacija ISO standardizirala koncepte, povezane s popolnim nadzorom nad kakovostjo, leta 1987 pa je nastal še danes pomemben standard ISO 9000 [4].

Čeprav se vsi avtorji strinjajo glede pomembnosti kakovosti, se njihove definicije kakovosti široko razlikujejo in tako v skupnosti ni soglasja o niti eni definiciji kakovosti. Tako je za nekatere kvaliteten proizvod proizvod brez napak, za druge pa proizvod, ki zadovoljuje uporabnikove potrebe. Večina avtorjev je tako definicije kakovosti razdelila v dve široki skupini: interna

kakovost (proizvodi brez napak) in eksterna kakovost (proizvodi, ki zadovoljujejo potrebe uporabnika).

Slika 4.1 predstavlja poenostavljen koncept interne in eksterne kakovosti. Na eni strani interna kakovost ustreza nivoju podobnosti med “popolnimi podatki”, ki bi jih lahko izdelali in podatki, ki so bili dejansko izdelani. Na drugi strani pa eksterna kakovost ustreza nivoju podobnosti med proizvedenimi podatki in uporabnikovimi zahtevami.



Slika 4.1: Poenostavljen koncept interne in eksterne kakovosti.

4.1 Interna kakovost

Interna kakovost ustreza nivoju podobnosti med izdelanimi podatki in popolnimi podatki, ki bi morali biti izdelani, se pravi podatki brez napak. Takšnim popolnim podatkom običajno rečemo “nominalna osnova”, ki jo lahko definiramo kot “sliko realnega sveta v danem času, skozi filter definiran s specifikacijami proizvoda” [2]. Specifikacije proizvoda so definirane kot “dokument, ki predpisuje zahteve katerim mora proizvod ustrezati” [2] in so niz pravil in zahtev, ki definirajo način prehoda iz realnega sveta v podatke. Vsebujejo, na primer definicije pojavov, ki naj bodo predstavljeni, geometrije, ki naj se jih uporabi za predstavitev posameznega tipa pojava, attribute, ki opisujejo pojave ter vse možne vrednosti teh atributov.

Nominalna osnova se v praksi za ocenjevanje interne kakovosti ne uporablja, saj ta, kot “idealni” niz podatkov, ne obstaja. Bolj primerni so podatkovni

nizi, ki so boljše kakovosti kot izdelani podatki, tako imenovani “kontrolni podatki” ali “referenčni podatki”. Ocena interne kakovosti bo tako vsebovala eksterno komponento (primerjavo z referenčnimi podatki) in interno komponento, odvisno od elementa kakovosti, ki se ocenjuje. Ocenjevanje natančnosti bo, na primer, izvedeno eksterno, medtem ko bo ocena topološke pravilnosti izvedena interno.

Interno kakovost lahko opišemo z različnimi kriteriji, za katere obstaja nek konsenz in jih uporabljajo tudi pomembnejši standardi na področju prostorske informatike (*geomatike*) (ISO, FGDC, CEN, ...). Standard o kakovosti prostorskih podatkov ISO 19113 [6] priporoča naslednje kriterije, oz. elemente kakovosti:

- **Popolnost** (angl. *completeness*) podaja prisotnost ali odsotnost pojavov, njihovih atributov in relacij v podatkovnem nizu. Iz perspektive ponudnika podatkov je merilo stopnje skladnosti podatkov z realnim svetom, gledanim skozi specifikacije za zajem podatkov. Delimo jo lahko na naslednja dva elementa:
 - Izostanki (angl. *omission*), kjer so entitete realnega sveta zajete v specifikacijah, vendar manjkajo v podatkovnem nizu.
 - Presežki (angl. *commission*), kjer pojavi ostajajo v podatkovnem nizu po tem, ko v realnem svetu ne več ali pa obstajajo v podatkih in se ne skladajo s specifikacijami za zajem podatkov.
- **Logična usklajenost** (angl. *logical consistency*) podaja stopnjo skladnosti med logičnimi ter pojmovnimi pravili podatkovnega modela in strukturo podatkov v podatkovnem nizu, kar se nanaša predvsem na sestavo razredov, atributov in relacij med njimi. Logična usklajenost se lahko deli na naslednje elemente:
 - Pomensko usklajenost (angl. *conceptual consistency*), ki podaja skladnost med pojmovnim modelom in njegovim formalnim opisom v pojmovni shemi.
 - Formatno usklajenost (angl. *format consistency*), ki podaja stopnjo usklajenosti shranjenih podatkov s fizično strukturo podatkovnega niza.
 - Domensko usklajenost (angl. *domain consistency*), ki podaja usklajenost podatkovnih vrednosti z zalogo vrednosti določenega atributa.

- Topološko usklajenost (angl. *topological consistency*), ki podaja pravilnost izrecno opredeljenih topoloških lastnosti gradnikov v podatkovnem nizu.
- **Položajna natančnost** (angl. *positional accuracy*) podaja točnost lege pojavov v podatkovnem nizu in se lahko deli na naslednje tri elemente:
 - Absolutno ali zunanjo natančnost (angl. *absolute or external accuracy*), ki podaja odstopanje koordinatnih leg v podatkovnem nizu od njihove stvarne lokacije.
 - Relativno ali notranjo natančnost (angl. *relative or internal accuracy*), ki podaja odstopanje relativno podanega položaja pojavov v podatkovnem nizu od njihove stvarne relativne lege.
 - Natančnost podatkov v mreži (angl. *gridded data position accuracy*), ki podaja odstopanje položaja podatkov v mreži od njihove stvarne lokacije.
- **Tematska natančnost** (angl. *thematic accuracy*) podaja zanesljivost izvedene klasifikacije pojavov ter točnost kvantitativnih in pravilnost kvalitativnih opisnih atributov v podatkovnem nizu. Delimo jo lahko na naslednje tri elemente:
 - Pravilnost klasifikacije (angl. *classification correctness*), ki podaja zanesljivost razvrščanja pojavov ali atributov v ustrezne razrede glede na stvarno stanje ali referenčne podatke.
 - Kvalitativno pravilnost (angl. *non-quantitative attribute correctness*), ki podaja doslednost neštevnih vrednosti opisnih atributov.
 - Kvantitativno točnost (angl. *quantitative attribute accuracy*), ki podaja zanesljivost števnih vrednosti opisnih atributov.
- **Časovna natančnost** (angl. *temporal accuracy*) podaja točnost časovnih atributov in časovnih odnosov med obravnavanimi pojavi v podatkovnem nizu in ima lahko naslednje tri elemente:
 - Točnost časovnih meritev (angl. *accuracy of time measurement*), ki podaja pravilnost in odstopanje časovnih podatkov o pojavih. Običajno je podana kot napaka časovne meritve.
 - Časovna usklajenost (angl. *temporal consistency*), ki podaja pravilnost razvrstitve časovnih pojavov, če je ta prisotna.

- Časovno veljavnost (angl. *temporal validity*), ki podaja veljavnost podatkov glede na čas.

4.2 Eksterna kakovost

Koncept eksterne kakovosti ustreza nivoju skladnosti med izdelanimi podatki in uporabniškimi zahtevami ali pričakovanji v danem kontekstu. Če bi več ljudi vprašali, kateri avtomobil, film ali jed je najboljša, bi dobilo veliko različnih odgovorov. Izbira vsake osebe bi bila odvisna od različnih kriterijev, preteklih izkušenj, osebnih preferenc in podobno. Če bi želeli primerjati dva avtomobila, na primer Škodo in Ferrarija, bi se pri oceni interne kakovosti verjetno bolje odrezal Ferrari. Glede na eksterno kakovost pa bi se lahko izkazal kot predrag, privlačen za tatove in neprimeren za vožnjo po slabših cestah. Za povprečnega potrošnika bi se tako Škoda izkazala za primernejšo in bi s tem imela boljšo eksterno kakovost. Podobno bi bil podatkovni niz hidrografije, z najboljšo interno kakovostjo in izdelan z najboljšimi metodami za nadzor nad kakovostjo, zelo uporaben za nekega okoljskega inženirja (dobra eksterna kakovost) in hkrati popolnoma neuporaben za geodeta (slaba eksterna kakovost).

Koncept eksterne kakovosti tako potrjuje, da kakovost ni absolutna in ima lahko isti izdelek različno kakovost za različne uporabnike. Ta koncept se običajno uporablja kot definicija kakovosti v širšem smislu. ISO [5] definira kakovost kot “Karakteristike izdelka, ki se nanašajo na njegovo zmožnost zadovoljiti navedene in naznačene potrebe”. Eksterna kakovost se zato pogosto definira kot “primernost za uporabo” ali “primernost namenu”.

Eksterno kakovost prostorskih podatkovnih nizov lahko definiramo z naslednjimi šestimi karakteristikami:

- **Definicija** ocenjuje ali točna narava podatkov in pojavov, ki jih opisuje, to je “kaj”, ustreza uporabnikovim potrebam (semantične, prostorske in časovne definicije).
- **Pokritost** ocenjuje ali območje in časovni razpon podatkov, to je “kje” in “kdaj”, ustrezata uporabnikovim potrebam.
- **Poreklo** opisuje od kje so podatki prišli, cilje njihove pridobitve, uporabljene metode za pridobitev in obdelavo podatkov, to je “kako” in “zakaj”.
- **Natančnost** ali preciznost ocenjuje, če so podatki dovolj natančni in s

tem sprejemljivi za uporabnikove potrebe (semantična, časovna in prostorska natančnost pojavov in njihovih atributov).

- **Legitimnost** ocenjuje ali so podatki uradno priznani, njihov pravni pomen, ustreznost *de facto* standardom, upoštevanje pomembnejših standardov, če jih pravno ali administrativno priznava kako uradno telo, če ponudnik podatkov nudi kako jamstvo in podobno.
- **Dostopnost** ocenjuje, na kakšen način lahko uporabniki dostopajo do podatkov (cena, časovni okvir, format, zaupnost, pravne omejitve ipd.)

Medtem ko metode za ocenjevanje interne kakovosti obstajajo, ocenjevanje eksterne kakovosti še vedno ostaja precej neraziskano področje. Zaradi opisne narave karakteristik eksterne kakovosti pa avtomatsko preverjanje le-te v večini primerov ni možno.

5 Metapodatki o kakovosti prostorskih podatkov

Prostorski metapodatki nudijo dodatne informacije o prostorskem podatkovnem nizu. Njihov namen je dokumentiranje opisanih podatkov. V splošnem zajemajo informacije o pomenu, vsebini, kakovosti, zgodovini, časovnosti, koordinatnem referenčnem sistemu, podatkovnem formatu, dostopnosti, vrednosti, pravnih in avtorskih omejitvah, organizaciji, ki je podatke izdelala, ter druge karakteristike prostorskih podatkov.

Metapodatki imajo pomembno vlogo pri izmenjavi prostorskih podatkov, na njihovi podlagi uporabniki lahko ocenijo, če so podatki primerni za njihov namen uporabe. Z zbiranjem metapodatkov v kataloge uporabnikom lahko omogočimo iskanje razpoložljivih podatkov.

Metapodatki lahko opisujejo podatke na različnih nivojih podrobnosti. ISO 19115 [8] predlaga naslednje nivoje: “podatkovni katalog”, “podatkovni niz”, “tip pojava”, “tip atributa”, “instanca pojava” in “instanca atributa”. Manj podrobni nivoji so opisani v metapodatkovni datoteki, bolj podrobni nivoji kot instanca pojava ali atributa pa kar kot dodatni atributi pojava. Denimo pri podatkih, kot so digitalni katastrski načrt in register geodetskih točk, mora vsaka točka imeti informacijo o datumu, metodi in napaki izmere.

Metapodatki so običajno zapisani v obliki dokumenta XML [9] in so v taki obliki najbolj pogost način za sporočanje informacij o kakovosti prostorskih podatkov. Različne organizacije objavljajo standarde za dokumentiranje kakovosti prostorskih podatkov v obliki metapodatkov (ISO 19113 [6] in ISO 19115 [8], direktiva INSPIRE [12, 13] in drugi). Ker pa so ti podatki sami običajno težko razumljivi uporabnikom in včasih celo ekspertom na področju geografskih informacij, je njihova uporabnost v takšni obliki omejena. Zato se vse več vlaga v razvoj in standardizacijo metod za grafično vizualizacijo teh podatkov [11].

Metapodatki, ki jih izdelajo ponudniki podatkov, so s strani uporabnikov pogosto dojeti kot zaščita ponudnikov pred potencialnimi tožbami namesto kot informacije v pomoč pri ocenjevanju primernosti za nameravano uporabo [3]. Zato je uporabnike nujno ozaveščati o pomembnosti in uporabnosti metapodatkov o kakovosti.

6 Standardi na področju kakovosti prostorskih podatkov

Da lahko primerjamo informacije o kakovosti različnih prostorskih podatkov iz različnih virov, potrebujemo standardiziran opis kakovosti, standardizirane metode ocenjevanja kakovosti in standardizirano obliko zapisa podatkov o kakovosti. Za Slovenijo in območje EU je pomembna predvsem direktiva INSPIRE ter mednarodni standardi, ki jih razvija ISO TC (*Technical Committee* - Tehnični odbor) 211 in jih je prevzel tudi Slovenski inštitut za standardizacijo (SIST).

6.1 ISO 19113:2002

Standard (*SIST EN*) *ISO 19113:2002 - Geografske informacije - Načela kakovosti* [6] podaja principe opisovanja kakovosti prostorskih podatkov, komponente za poročanje o kakovosti in predlaga pristop k organiziranju informacij o kakovosti. Kakovost podatkovnega niza opisuje s kriteriji oziroma elementi, ki smo jih opisali v poglavju 4.1 ter s preglednimi elementi kakovosti, ki podajajo splošne, nekvantitativne informacije (namen, poreklo in način uporabe). Ta standard ne skuša definirati minimalne sprejemljive kakovosti prostorskih podatkov.

6.2 ISO 19114:2003

Standard (*SIST EN*) *ISO 19114:2003 - Geografske informacije - Postopki za ocenjevanje kakovosti* [7] predpisuje metodologijo za določanje in ocenjevanje kvantitativne kakovosti prostorskih podatkov v skladu s standardom ISO 19113 ter podaja napotke za izdelavo in sestavo poročila o rezultatih ocenjevanja, ki so lahko v obliki metapodatkov, skladnih s standardom ISO 19115 ali v drugi obliki standardnega poročila.

Standard opisuje proces ocenjevanja kakovosti kot zaporedje korakov za izdelavo in sporočanje podatkov o kakovosti. Opisani proces zajema določanje obsega testiranja, izbiro primernih elementov kakovosti, izbiro mere kakovosti, izbiro in uporabo metod za ocenjevanje kakovosti, izdelavo poročila in opsijski korak določanja skladnosti kakovosti s specifikacijami podatkov. Ta standard predpisuje le smernice za implementacijo metod ocenjevanja kakovosti in ne definira nobene konkretne metode.

6.3 ISO 19115:2003

Standard (*SIST EN*) *ISO 19115:2003 - Geografske informacije - Metapodatki* [8] predpisuje uporabo metapodatkov in v jeziku UML formalizira koncepte kakovosti iz standarda ISO 19113 in obliko poročila o rezultatih ocenjevanja skladno z metodologijo standarda ISO 19114. Standard je organiziran v metapodatkovne sklope, med katerimi je en sklop namenjen posebej informacijam o kakovosti.

Določa obvezne in opsijske metapodatkovne sekcije, entitete in elemente, minimalni set metapodatkov, ki so potrebni za uporabo v aplikacijah širokega obsega (raziskovanje podatkov, ugotavljanje primernosti za nameravano uporabo, dostop do podatkov, prenos podatkov ter način uporabe podatkov), ter standardiziran način razširitve metapodatkov za specializirane potrebe.

Zaradi obsežnosti (vsebuje preko 400 metapodatkovnih elementov v 15 sekcijah) ga uporabniki pogosto ne razumejo, kar vodi v nepravilno rabo metapodatkovnih informacij. Kljub obsežnosti pa ne pokriva vseh potreb, med drugim ne vsebuje elementov za opis metode vzorčenja in ga je pogosto potrebno razširjati, kar povzroča še dodatno nerazumljivost.

6.4 ISO 19139:2007

Ker ISO 19115 ne določa oblike zapisa metapodatkov, je bila implementacija metapodatkov odvisna od posameznih ponudnikov podatkov. Da bi omogočili standardizacijo implementacij, je bil sprejet standard (*SIST-TS*) *ISO 19139 - Geografske informacije - Metapodatki - Implementacija sheme XML* [9], ki definira shemo XML za zapis metapodatkov iz standarda ISO 19115 v obliki dokumentov XML in tako nudi enotno specifikacijo za opisovanje, izmenjavo in preverjanje veljavnosti metapodatkov.

6.5 INSPIRE

Direktivo INSPIRE [12] je sprejela Evropska unija 14.3.2007 z namenom, da bi omogočila skupno in primerljivo infrastrukturo za izmenjavo prostorskih podatkov znotraj EU za namen spremljanja regulativ, ki imajo vpliv na okolje.

INSPIRE ne predpisuje kakovosti, ki bi ji morali zadoščati prostorski podatki držav članic, ampak metodologijo merjenja le-te ter uporabo orodij za širjenje podatkov in informacij o samih podatkih, vključno s kakovostjo - uporaba metapodatkov, mrežnih storitev, nudenje podatkov in storitev ter nadzor in poročanje. INSPIRE temelji na prevzetih in prilagojenih standardih ISO.

Na problematiko obravnave kakovosti prostorskih podatkov nakazuje dejstvo, da je od 80 objav na JRC Wiki (portal znanja *Joint Research Centra*), ki se nanašajo na rezultate testiranja, samo 9 udeležencev preizkusilo kakovost podatkov in samo 5 od teh tako kakovost kot tudi metapodatke. Za dejavnike, ki so vodili v tako stanje, so bili navedeni pomanjkanje ustreznih orodij na trgu, pomanjkanje časa in sredstev ter potreba po izobraževanju na področju tolmačenja INSPIRE specifikacij [14].

Šele v tem obdobju so se začeli praktični poskusi, da bi v okviru INSPIRE natančno definirali kakovost. ESDIN (*European Spatial Data Infrastructure best practice Network*) in podjetje 1Spatial skupaj razvijata t.i. konformni model prostorske podatkovne infrastrukture (*SDI Conformance Model*).

Navkljub temu, da je Evropska unija uzakonila direktivo INSPIRE ter predpisala metodologije ravnanja s prostorskimi podatki, bo glede na trenutno stanje za praktično uporabo teh podatkov ustrezne kakovosti potrebno počakati še nekaj časa.

7 Metode za avtomatsko ocenjevanje interne kakovosti

Za ocenjevanje kakovosti podatkovnega niza morajo obstajati jasno opredeljene metode oziroma postopki, ki jih je treba dosledno uporabljati. Tak pristop proizvajalcem podatkov omogoča, da lahko objektivno izrazijo skladnost samih podatkov z njihovimi specifikacijami. Uporabniki pa lahko ocenijo, če podatki ustrezajo njihovim zahtevam.

Ocenjevanje kakovosti se lahko izvaja v različnih fazah življenjskega cikla podatkovnega niza. Proizvajalci kakovost običajno ocenijo že nad vzorčnimi podatki med pripravo specifikacije podatkovnega niza, da preprečijo morebitne neskladnosti modela s specifikacijami in zaznajo morebitne probleme z uporabljenimi metodologijo zajema že pred dejanskim zajemom podatkov. Proizvajalec mora ocenjevanje izvesti tudi na končnem izdelku in vsaki posodobljeni verziji ter izdelati ustrezno poročilo o kakovosti.

Če uporabniki potrebujejo več informacij o kakovosti podatkovnega niza, kot jih je podal proizvajalec ali če metode, ki jih je uporabil proizvajalec niso ustrezne za oceno eksterne kakovosti pri nameravani uporabi, lahko uporabniki preverijo ali ponovijo proizvajalčeve metode, oziroma uporabijo druge metode.

Metode za ocenjevanje interne kakovosti podatkovnih nizov se delijo na [8]:

- **neposredne** (*direktne*) metode, ki določajo kakovost s primerjavo podatkov z notranjimi ali z zunanji referenčnimi informacijami. Glede na vir informacij jih naprej delimo na:
 - **notranje** metode, kjer so vsi podatki, potrebni za oceno kakovosti, vsebovani v podatkovnem nizu samem. Na primer: za oceno topološke usklajenosti, so vsi potrebni podatki nahajajo v sami topološki strukturi.
 - **zunanje** metode, kjer so za oceno kakovosti potrebni dodatni referenčni podatki, ki se nahajajo izven ocenjevanega podatkovnega niza. Na primer: za oceno položajne natančnosti potrebujemo referenčni podatkovni niz ali dodatne oziroma nove meritve.
- **posredne** (*indirektne*) metode, ki ocenjujejo kakovost glede na poznane zunanje informacije. Take informacije običajno nudijo elementi eksterne kakovosti, druga poročila o kakovosti ali izvorni podatki. Rezul-

tati posrednih metod so opisne narave, njihova avtomatizacija zato ni mogoča. Uporabljajo se predvsem v primerih, ko uporaba neposrednih metod ni možna.

Metode neposrednega ocenjevanja lahko izvedemo ročno ali avtomatsko. Če razpolagamo z ustreznimi referenčnimi podatki, lahko z uporabo notranjih in zunanjih metod avtomatsko ocenimo vse elemente interne kakovosti. Samo z uporabo notranjih metod pa lahko preverimo le formatno, domensko, časovno, pomensko in topološko usklajenost.

Kakovost lahko ocenjujemo na celotnem podatkovnem nizu ali na vzorcu, določenim s primerno metodo vzorčenja prostorskih podatkov [8].

ISO 19114 [8] in sorodni standardi predpisujejo le splošne smernice in metodologije za ocenjevanje kakovosti. Definiranje in uporaba konkretnih metod pa je običajno stvar posameznih organizacij. Izbira metod je odvisna tudi od tipa in lastnosti posameznega podatkovnega niza. Tako je nastalo več različnih metod za ocenjevanje kakovosti, v poročilu o kakovosti pa mora biti navedeno, katere od teh metod so bile uporabljene za oceno posameznega elementa kakovosti.

V nadaljevanju predstavljene metode za avtomatsko ocenjevanje posameznih elementov kakovosti se nanašajo le na vektorske podatke. Za razumevanje metod ocenjevanja kakovosti rastrskih podatkov je potrebno znanje iz področja radiometrije ter podrobno poznavanje delovanja opreme in procesa za zajem rastrskih podatkov [2], kar pa je izven obsega tega diplomskega dela.

7.1 Formatna usklajenost

Pri ocenjevanju kakovosti običajno najprej preverimo formatno usklajenost. Napake v fizični strukturi podatkovnega niza ali neskladnosti formata podatkov s specifikacijo podatkovnega niza namreč lahko vodijo v nedelovanje ali nepredvidljivo delovanje metod za ocenjevanje drugih elementov kakovosti. Podatkovne nize s neusklajenim formatom običajno zavrnamo kot neustrezne.

Format vseh geometrijskih objektov najlažje preverimo, če iz datoteke podatkovnega niza poskušamo prebrati vsak objekt. Programska koda za branje pričakuje ustrezen format zapisa geometrij in bo ob napačnem formatu vrnila napako. Če je bilo branje uspešno, lahko zapise geometrijskih objektov pretvo-

rino v notranji format programske opreme in dodatno preverimo njihovo logično pravilnost.

Lastnosti polj atributnih podatkov moramo primerjati s specifikacijami podatkovnega niza. Preveriti moramo prisotnost vseh polj, njihov tip, velikost, edinstvenost in nepraznost. Če definicija tipa in strukture polja ne zagotavlja predpisanega formata, moram preveriti format vsakega zapisa v podatkovnem nizu.

Implementacija metod za preverjanje formatne usklajenosti je odvisna od tipa datoteke ali podatkovnega vira v katerem je shranjen podatkovni niz. Na primer v podatkovni bazi SQL lahko vse našteto preverimo s stavki SQL.

7.2 Domenska in časovna usklajenost

Domeno ali zalogo vrednosti atributnih polj lahko določimo na dva načina: kot razpon možnih vrednosti ali s šifrantom vrednosti. Vzemimo primer katastrskih parcel, kjer polju za zapis uradne površine v kvadratnih metrih lahko določimo, da morajo biti vse vrednosti v tem polju večje od 10, vrednosti v polju za zapis parcelne številke morajo biti med 1 in 10000, vrednosti polja za zapis vrste parcele pa morajo biti prisotne v šifrantu vrst parcel.

Ker je časovna komponenta zapisov običajno shranjena kar kot atributni podatek, lahko časovno usklajenost (kadar je ta določena kot dovoljen časovni razpon) obravnavamo enako kot domensko usklajenost.

Podobno kot pri formatni usklajenosti, je tudi implementacija metod za preverjanje domenske usklajenosti odvisna od tipa datoteke oziroma podatkovnega vira v katerem je shranjen podatkovni niz. V podatkovni bazi SQL lahko usklajenost vrednosti z domeno preverimo s preprostimi stavki SQL. V drugih podatkovnih virih pa je običajno potrebno preveriti vrednosti vsakega zapisa posebej.

7.3 Presežki in izostanki ter tematska natančnost

Presežki in izostanki ter tematska natančnost so merila skladnosti podatkov o pojavih z realnim svetom, gledanim skozi specifikacije za zajem podatkov. Ocenjevanje teh elementov kakovosti lahko izvedemo le z uporabo zunanjih

metod, torej potrebujemo referenčni podatkovni niz, zajet z združljivimi specifikacijami v čim bolj istem času ali pa dodatne oziroma nove meritve.

Ker so primerni referenčni podatki le redko razpoložljivi, se ocenjevanje običajno izvaja na dodatnih meritvah boljše tematske natančnosti ali zanesljivosti. Z uporabo teh meritev pa je ocenjevanje omejeno le na vzorec, ki ga zajemajo meritve, ocena celotnega podatkovnega niza tako ni mogoča.

Če med ocenjevanim podatkovnim nizom in referenčnimi podatki (sem spadajo tudi nove meritve) obstaja povezovalni ključ, je določanje izostankov in presežkov podobno dvostranskemu preverjanju domenske usklajenosti s šifrantom vrednosti. Za referenčne podatke v tem primeru lahko uporabimo tudi neprostorske podatkovne nize. Denimo imena cest v prostorskem podatkovnem nizu cest lahko primerjamo z nekim registrom imen cest.

Ko povezovanje zapisov preko povezovalnega ključa ni možno, je zapise potrebno povezati preko njihove prostorske lokacije. Za takšno avtomatsko povezovanje morajo tako ocenjevani kot referenčni podatki imeti zadostno položajno natančnost. V nasprotnem primeru lahko povežemo napačne zapise. Ker je takšno položajno natančnost težko zagotoviti, se za povezovanje zapisov preko njihove lokacije pogosto uporablja ročne metode.

Ko najdemo ustrezen referenčen zapis (če ta obstaja), lahko tematsko natančnost preverimo s preprostim primerjanjem vrednosti ocenjevanih atributov. Kriteriji za določitev natančnosti so različni, lahko zahtevamo popolno uje-manje vrednosti ali pa dovolimo določeno odstopanje.

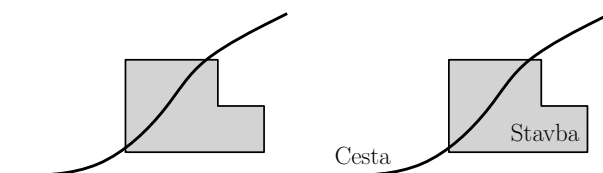
7.4 Pomenska usklajenost ali integriteta prostorskih podatkov

Pomenska usklajenost ali integriteta prostorskih podatkov je pogosto zanemarjen aspekt kakovosti prostorskih podatkov, ki se mu ne posveča dovolj pozornosti. Običajno jo delimo v dve kategoriji: metrične relacije in topološke relacije. Metrične relacije se ukvarjajo z merljivimi karakteristikami pojavov, kot so na primer prostorska in časovna razdalja ter usmerjenost. Topološke relacije pa se ukvarjajo z geometrijskimi lastnostmi, kot so prekrivanja in nepovezanosti.

V tem poglavju pogledjmo kako z uporabo topoloških relacij zagotovimo in-

tegriteto prostorskih podatkov in tako izboljšamo njihovo interno kakovost. Najprej vzemimo primer topološke karte, kjer bi poleg položajne natančnosti in topološke pravilnosti geometrij želeli, da ceste ne prečkajo stavb in vodnih površin, da so otoki znotraj vodnih površin, da so mostovi in predori povezani s cesto, potjo ali železnico, da stavbe ne ležijo v jezerih ipd. Takšno zeleno stanje podatkov lahko opišemo z integritetnimi omejitvami prostorskih podatkov.

Slika 7.1 predstavlja primer geometrijskega protislovja med cesto in stavbo, dvema objektoma distinktivnega pomena. V splošnem bi linija lahko prečkala poligon (levo), kar ne bi bilo protislovno, dokler na primer linija ne predstavlja ceste in poligon stavbe (desno).



Slika 7.1: Primer pomenske neusklajenosti med cesto in stavbo.

Takšna protislovja se lahko pojavijo ob vnosu podatkov (na primer napačna klasifikacija objekta ali generalizacija objektov) ali po integraciji s podatkovnim nizom iz drugega vira ali s podatkovnim nizom drugačne natančnosti. S preverjanjem veljavnosti prostorskih integritetnih omejitev lahko zmanjšamo vnos takšnih napak in zaznamo protislovja v integriranih podatkih.







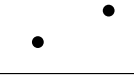
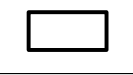



7.4.1 Prostorski relacijski operatorji

Prostorski operatorji so funkcije z vsaj enim prostorskim parametrom. Prostorski relacijski operatorji pa so Boolove metode za preverjanje specifičnih topoloških prostorskih relacij med dvema geometrijskima objektoma [10]. Osnovni princip za primerjanje dveh geometrijskih objektov je parno testiranje presekov med notranjostma, mejama in zunanostma objektov (slika 7.2) in klasifikacija relacij na podlagi ugotovljenih presekov.

Meja geometrijskega objekta je množica geometrijskih objektov naslednje nižje dimenzije. Meja točke ali zbirke točk je torej prazna množica. Meja zaključene krivulje je prav tako prazna množica, meje nezaključene krivulje pa je sestavljena iz njenih dveh končnih točk. Mejo zbirke krivulj sestavljajo tiste točke, ki so na meji lihega števila njenih elementarnih krivulj. Meja poligona je

množica njegovih zaprtih krivulj - obročev, meja zbirke poligonov pa množica obročev vseh poligonov. Mejo poljubne zbirke geometrijskih objektov disjunkt-nih notranjosti sestavljajo geometrijski objekti, dobljeni iz unije njihovih mej po modulu 2 [10].

Notranjost geometrijskega objekta sestavljajo točke, ki ostanejo po odstranitvi mejnih točk. Zunanost geometrijskega objekta sestavljajo točke, ki niso ne na meji in ne v notranjosti objekta.

	Točka	Linija	Poligon
			
Notranjost (N)			
Meja (M)			
Zunanost (Z)			

Slika 7.2: Koncept notranjosti, meje in zunanosti geometrijskega objekta.

Da bi formalizirali prostorske relacije med geometrijskimi objekti, je bil sprva predlagan 4-presečni model za primerjavo mej in notranjosti med dvema objektoma, ki se ga je z upoštevanjem preseka z zunanostmi objektov razširilo v 9-presečni model. Nadaljnje razširjen model, ki ga predlaga standard *OGC Simple Feature Access* [10], vključuje tudi informacijo o dimenziji preseka, "Dimenzijsko razširjen 9-presečni model" s formalno oznako *DE-9IM*.

7.4.1.1 Dimenzijsko razširjen 9-presečni model (DE-9IM)

Naj $N(a)$, $M(a)$ in $Z(a)$ predstavljajo notranjost, mejo in zunanost poljubnega geometrijskega objekta a .

Funkcija $\dim(x)$ naj vrne najvišjo dimenzijo (-1, 0, 1 ali 2) geometrijskega objekta x , kjer numerična vrednost -1 predstavlja prazno množico oz. $\dim(\emptyset)$.

Presek poljubnih dveh $N(a)$, $M(a)$ ali $Z(a)$ je lahko množica geometrijskih

objektov x z različnimi dimenzijami (tabela 7.1). Presek mej dveh poligonov na primer lahko vsebuje linijo in točko.

Tabela 7.1: Osnovna oblika modela DE-9IM [10].

	Notrajnost (b)	Meja (b)	Zunanost (b)
Notranost (a)	$\dim(N(a) \cap N(b))$	$\dim(N(a) \cap M(b))$	$\dim(N(a) \cap Z(b))$
Meja (a)	$\dim(M(a) \cap N(b))$	$\dim(M(a) \cap M(b))$	$\dim(M(a) \cap Z(b))$
Zunanost (a)	$\dim(Z(a) \cap N(b))$	$\dim(Z(a) \cap M(b))$	$\dim(Z(a) \cap Z(b))$

Prostorski relacijski predikat za dva geometrijska objekta lahko izrazimo kot formulo, ki za vhod sprejme vzorčno matriko, ki predstavlja množico sprejemljivih vrednosti v presečnem modelu. Predikat je resničen, če prostorska relacija med objektoma ustreza eni od sprejemljivih vrednosti v vzorčni matriki, ki sestoji iz množice devetih vzorčnih vrednosti, po eno za vsako celico v presečnem modelu. Vzorčne vrednosti p lahko zavzemajo vrednosti $\{P, N, *, 0, 1, 2\}$, ki so definirane takole:

$$\begin{aligned}
 p = P &\Rightarrow \dim(x) \in \{0, 1, 2\}; x \neq \emptyset \\
 p = N &\Rightarrow \dim(x) = -1; x \neq \emptyset \\
 p = * &\Rightarrow \dim(x) \in \{-1, 0, 1, 2\}; \text{ poljubna vrednost} \\
 p = 0 &\Rightarrow \dim(x) = 0 \\
 p = 1 &\Rightarrow \dim(x) = 1 \\
 p = 2 &\Rightarrow \dim(x) = 2
 \end{aligned}$$

Vzorčno matriko lahko zapišemo kot niz devetih znakov, npr. "P*P***P**". Za primer podajmo del psevdokode za preverjanje prekrivanja med dvema območjema:

*Niz matrikaPrekrivanja = "P*P***P**"*
Geometrija a, b
Boolean b = a. vRelaciji(b, matrikaPrekrivanja)

7.4.1.2 Poimenovani predikati prostorskih relacij v modelu DE-9IM

Relacijski predikati na osnovi vzorčne matrike imajo prednost, da omogočajo testiranje velikega števila prostorskih relacij, a jih kot nizko nivojski gradnik težko opišemo v naravnem jeziku. Za prikladnejšo uporabo v visokonivojskih programskih jezikih in v poizvedbah SQL obstaja nekaj tako imenovanih “poimenovanih predikatov”. Standard *OGC Simple Feature Access* [10] definira osem poimenovanih predikatov, razloženih v nadaljevanju.

Za kompaktnejši zapis definicij bomo uporabili naslednjo notacijo za opis skupin geometrijskih objektov: P predstavlja 0-dimenzijske geometrije (točke in zbirke točk), L 1-dimenzijske geometrije (krivulje in zbirke krivulj) in A 2-dimenzijske geometrije (poligoni in zbirke poligonov).

Enak (angl. *Equals*)

Dva topološko zaprta geometrijska objekta a in b imata enako notranjost in mejo. Velja:

$$a.Enak(b) \Leftrightarrow a \subseteq b \wedge b \subseteq a, \quad (1)$$

oziroma v terminologiji modela DE-9IM:

$$\begin{aligned} a.Enak(b) &\Leftrightarrow (N(a) \cap N(b) \neq \emptyset) \wedge \\ &(N(a) \cap M(b) = \emptyset) \wedge \\ &(N(a) \cap Z(b) = \emptyset) \wedge \\ &(M(a) \cap N(b) = \emptyset) \wedge \\ &(M(a) \cap M(b) \neq \emptyset) \wedge \\ &(M(a) \cap Z(b) = \emptyset) \wedge \\ &(Z(a) \cap N(b) = \emptyset) \wedge \\ &(Z(a) \cap M(b) = \emptyset) \wedge \\ &(Z(a) \cap Z(b) \neq \emptyset) \\ &\Leftrightarrow a.vRelaciji(b, \text{“PNNNPNNNP”}) \end{aligned} \quad (2)$$

Nepovezan (angl. *Disjoint*)

Meje in notranjosti dveh topološko zaprtih geometrijskih objektov a in b se ne sekajo. Velja:

$$a.Nepovezan(b) \Leftrightarrow a \cap b = \emptyset, \quad (3)$$

oziroma v terminologiji modela DE-9IM:

$$\begin{aligned}
 a.Nepovezan(b) &\Leftrightarrow (N(a) \cap N(b) = \emptyset) \wedge \\
 &\quad (N(a) \cap M(b) = \emptyset) \wedge \\
 &\quad (M(a) \cap N(b) = \emptyset) \wedge \\
 &\quad (M(a) \cap M(b) = \emptyset) \\
 &\Leftrightarrow a.vRelaciji(b, "NN * NN * * * *").
 \end{aligned} \tag{4}$$

Se dotika (angl. *Touches*)

Meje dveh geometrijskih objektov a in b se sekajo, notranjosti pa ne. Relacija je definirana le nad skupinami objektov A/A, L/L, L/A, P/A in P/L, med P/P pa ne. Se pravi relacija ne velja med točkami in zbirkami točk. Velja:

$$a.SeDotika(b) \Leftrightarrow (N(a) \cap N(b) = \emptyset) \wedge (a \cap b) \neq \emptyset, \tag{5}$$

oziroma v terminologiji modela DE-9IM:

$$\begin{aligned}
 a.SeDotika(b) &\Leftrightarrow (N(a) \cap N(b) = \emptyset) \wedge \\
 &\quad ((M(a) \cap N(b) \neq \emptyset) \vee \\
 &\quad (N(a) \cap M(b) \neq \emptyset) \vee \\
 &\quad (M(a) \cap M(b) \neq \emptyset)) \\
 &\Leftrightarrow a.vRelaciji(b, "NP * * * * * *") \vee \\
 &\quad a.vRelaciji(b, "N * * P * * * * *") \vee \\
 &\quad a.vRelaciji(b, "N * * * P * * * *").
 \end{aligned} \tag{6}$$

Križa (angl. *Crosses*)

Presek notranjosti geometrijskih objektov a in b je geometrijski objekt nižje dimenzije, objekta pa se ne smeta vsebovati. Relacija je definirana le nad skupinam objektov P/L, P/A, L/L in L/A. Velja:

$$\begin{aligned}
 a.Križa(b) &\Leftrightarrow \dim(N(a) \cap N(b)) < \max(\dim(N(a)), \dim(N(b))) \wedge \\
 &\quad (a \cap b \neq a) \wedge (a \cap b \neq b),
 \end{aligned} \tag{7}$$

oziroma v terminologiji modela DE-9IM:

ko $a \in P, b \in L$ ali

ko $a \in P, b \in A$ ali

ko $a \in L, b \in A$:

$$\begin{aligned} a.Križa(b) &\Leftrightarrow (N(a) \cap N(b) \neq \emptyset) \wedge (N(a) \cap Z(b) \neq \emptyset) \\ &\Leftrightarrow a.vRelaciji(b, "P * P * * * * *") \end{aligned} \quad (8)$$

ko $a \in L, b \in L$:

$$\begin{aligned} a.Križa(b) &\Leftrightarrow \dim(N(a) \cap N(b)) = 0 \\ &\Leftrightarrow a.vRelaciji(b, "0 * * * * * *"). \end{aligned}$$

Znotraj (angl. *Within*)

Notranjost geometrijskega objekta a je deloma znotraj notranjosti geometrijskega objekta b . Velja:

$$a.Znotraj(b) \Leftrightarrow (a \cap b = a) \wedge (N(a) \cap Z(b) = \emptyset), \quad (9)$$

oziroma v terminologiji modela DE-9IM:

$$\begin{aligned} a.Znotraj(b) &\Leftrightarrow (N(a) \cap N(b) \neq \emptyset) \wedge (N(a) \cap Z(b) = \emptyset) \wedge (M(a) \cap Z(b) = \emptyset) \\ &\Leftrightarrow a.vRelaciji(b, "P * N * * N * * *"). \end{aligned} \quad (10)$$

Prekriva (angl. *Overlaps*)

Notranjost geometrijskega objekta a je deloma znotraj notranjosti geometrijskega objekta b . Velja:

$$\begin{aligned} a.Prekriva(b) &\Leftrightarrow \dim(N(a)) = \dim(N(b)) = \dim(N(a) \cap N(b)) \wedge \\ &(a \cap b \neq a) \wedge (a \cap b \neq b), \end{aligned} \quad (11)$$

oziroma v terminologiji modela DE-9IM:

$$\begin{aligned}
& \text{ko } a \in P, b \in P \text{ ali} \\
& \text{ko } a \in A, b \in A : \\
& a.Prekriva(b) \Leftrightarrow (N(a) \cap N(b) \neq \emptyset) \wedge \\
& \quad (N(a) \cap Z(b) \neq \emptyset) \wedge \\
& \quad (Z(a) \cap N(b) \neq \emptyset) \\
& \Leftrightarrow a.vRelaciji(b, "P * P * * * P * *"), \tag{12}
\end{aligned}$$

$$\begin{aligned}
& \text{ko } a \in L, b \in L : \\
& a.Prekriva(b) \Leftrightarrow (\dim(N(a) \cap N(b)) = 1) \wedge \\
& \quad (N(a) \cap Z(b) \neq \emptyset) \wedge \\
& \quad (Z(a) \cap N(b) \neq \emptyset) \\
& \Leftrightarrow a.vRelaciji(b, "1 * P * * * P * *").
\end{aligned}$$

Zadnja dva predikata sta izvedena in definirana z namenom prikladnejše uporabe:

Vsebuje (angl. *Contains*)

$$a.Vsebuje(b) \Leftrightarrow b.Znotraj(a) \tag{13}$$

Seka (angl. *Intersects*)

$$a.Seka(b) \Leftrightarrow !a.Nepovezan(b) \tag{14}$$

Kot vidimo, je za preverjanje prostorskih relacij potrebno le iskanje presekov med geometrijami in klasificiranje relacije na podlagi ugotovljenega preseka. Algoritem za iskanje presekov si bomo podrobneje pogledali v razdelku 7.5.2.2.

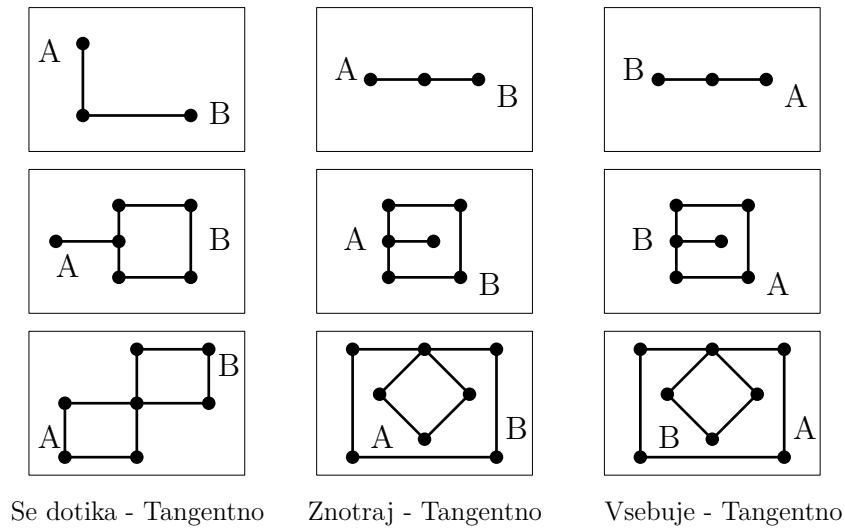
7.4.1.3 Razširitve predikatov prostorskih relacij

Predikati prostorskih relacij, definirani v standardu *OGC Simple Feature Access* [10] omogočajo opis posplošenih relacij, ki ne zadostujejo opisu nekaterih bolj specifičnih primerov. Zato se za natančnejši opis relacij pogosto uporablja razširitve *tangentno, meji na in striktno* [2]:

Razširitev tangentno (angl. *Tangent extension*)

Rezultat preseka mej geometrijskih objektov a in b mora biti 0-dimenzijski geometrijski objekt. Kar pomeni, da se morata objekta sekati le v eni ali več točkah. Relacija je definirana le nad skupinami objektov L/L, L/A, A/L ter A/A in jo lahko združujemo z relacijami *se dotika*, *znotraj in vsebuje* (slika 7.3). Velja:

$$\begin{aligned}
 a.SeDotika(b) \wedge a.Tangenten(b) &\Leftrightarrow a.SeDotika(b) \wedge (\dim(M(a) \cap M(b)) = 0) \\
 a.Znotraj(b) \wedge a.Tangenten(b) &\Leftrightarrow a.Znotraj(b) \wedge (\dim(M(a) \cap M(b)) = 0) \\
 a.Vsebuje(b) \wedge a.Tangenten(b) &\Leftrightarrow a.Vsebuje(b) \wedge (\dim(M(a) \cap M(b)) = 0)
 \end{aligned}
 \tag{15}$$



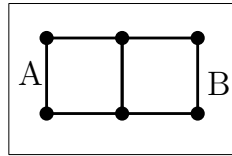
Slika 7.3: Razširitev *tangentno*.

Razširitev meji na (angl. *Borders extension*)

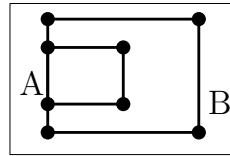
Rezultat preseka mej poligonov ali zbirk poligonov a in b mora biti 1-dimenzijski geometrijski objekt. Iz definicije sledi, da je relacija definirana le za skupino objektov A/A. Lahko jo združujemo z relacijami *se dotika*, *znotraj in vsebuje*

(slika 7.4). Velja:

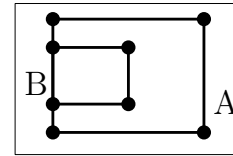
$$\begin{aligned}
 a.SeDotika(b) \wedge a.MejiNa(b) &\Leftrightarrow a.SeDotika(b) \wedge (\dim(M(a) \cap M(b)) = 1) \\
 a.Znotraj(b) \wedge a.MejiNa(b) &\Leftrightarrow a.Znotraj(b) \wedge (\dim(M(a) \cap M(b)) = 1) \\
 a.Vsebuje(b) \wedge a.MejiNa(b) &\Leftrightarrow a.Vsebuje(b) \wedge (\dim(M(a) \cap M(b)) = 1)
 \end{aligned}
 \tag{16}$$



Se dotika - Meji na



Znotraj - Meji na



Vsebuje - Meji na

Slika 7.4: Razširitev *meji na*.

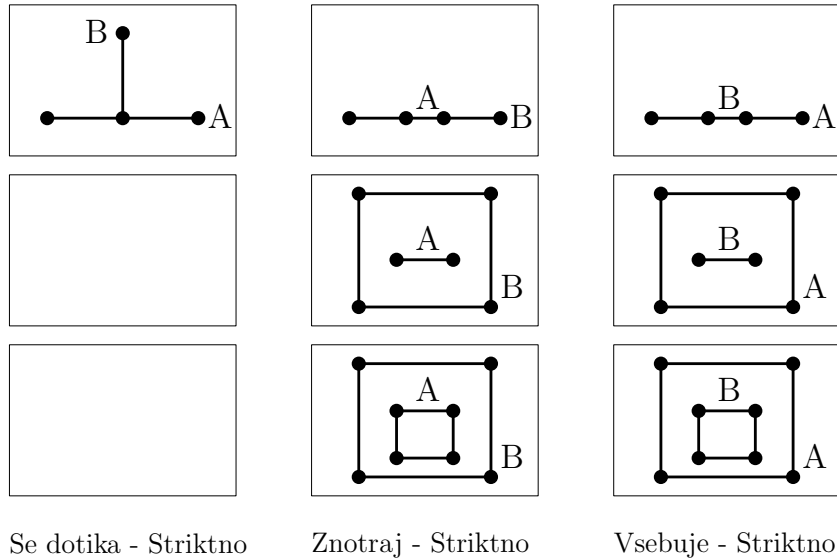
Razširitev striktno (angl. *Strict extension*)

Rezultat preseka mej geometrijskih objektov a in b mora biti prazna množica. Relacija je definirana le nad skupinami objektov L/L , L/A , A/L in A/A (slika 7.5). Velja:

$$\begin{aligned}
 a.SeDotika(b) \wedge a.Striktno(b) &\Leftrightarrow a.SeDotika(b) \wedge (\dim(M(a) \cap M(b)) = \emptyset) \\
 a.Znotraj(b) \wedge a.Striktno(b) &\Leftrightarrow a.Znotraj(b) \wedge (\dim(M(a) \cap M(b)) = \emptyset) \\
 a.Vsebuje(b) \wedge a.Striktno(b) &\Leftrightarrow a.Vsebuje(b) \wedge (\dim(M(a) \cap M(b)) = \emptyset)
 \end{aligned}
 \tag{17}$$

7.4.2 Kardinalnost

Prostorski operator po definiciji primerja par prostorskih objektov in kot rezultat vrne logično vrednost. Kardinalnost pove, s koliko objekti je nek objekt lahko v relaciji. Zapišemo jo s parom celih števil, ki določa največje in najmanjše število specifičnih relacij med dvema objektoma. Na primer krivulja, ki predstavlja železnico se mora dotikati krivulje, ki predstavlja cesto, kar zagotavlja, da nobena železnica ni izolirana od cestne infrastrukture. V realnosti bi pričakovali, da je vsaka železnica povezana z vsaj dvema cestama, po eno na vsakem koncu poti. Integritetni omejitvi *železnica.seDotika(Cesta)* bi tako dodali kardinalnost $(2, n)$.



Slika 7.5: Razširitev *striktno*.

7.4.3 Pas evklidske oddaljenosti

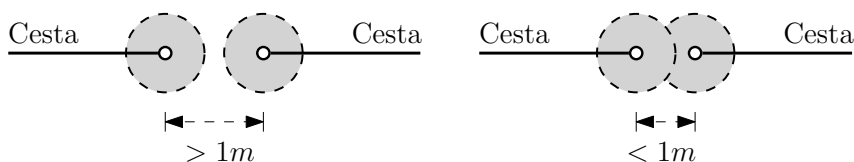
Pri delu s prostorskimi operatorji pogosto naletimo na specifična problema preciznosti koordinatnih meritev in prostorske bližine objektov v realnem svetu.

Algoritmi prostorskih operatorjev točke ali vozlišča običajno obravnavajo kot enake, le če imajo identične koordinate. V realnosti večina programskih rešitev pri procesiranju prostorskih podatkov uporablja različne nivoje natančnosti (število signifikantnih decimalnih mest) in včasih tudi različne podatkovne tipe za zapis koordinat (cela ali realna števila). To seveda otežuje uporabo prostorskih operatorjev, še posebej pri delu s podatki iz neenotnih virov. Problem prostorske bližine objektov v realnem svetu lahko ponazorimo s primerom ceste in železnice, ki se v realnosti včasih ne stikata popolnoma, a vendar jih upoštevamo kot povezani, če sta bližje kot, recimo, 100 metrov.

Nekatere programske rešitve te probleme rešujejo s spajanjem koordinatnih točk na presečišča celic drobno razdeljene mreže. Ta pristop zaokrožuje koordinate iz vseh podatkovnih virov na skupno referenco in omogoča boljšo natančnost pri iskanju integritetnih napak v prostorskih podatkih. To pa pomeni, da se integriteta podatkov preverja na virtualni verziji podatkov z zaokroženimi koordinatami namesto na originalnih podatkih.

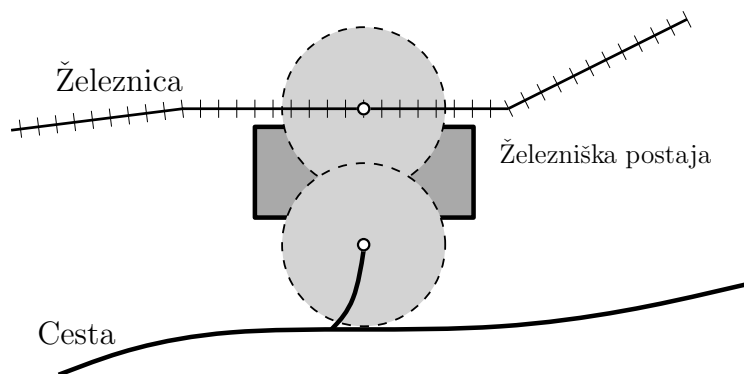
Boljši je pristop s tako imenovanim “pasom evklidske oddaljenosti” (angl. *buffer*), ki si ga lahko predstavljamo z okrog točke očrtano krožnico z določenim radijem. Točki imamo tako za enaki, če se njuni krožnici sekata oziroma dotikata.

Slika 7.6 ponazarja primer ulic, ki jim na končne točke očrtamo krožnice s polmerom pol metra. Dve ulici imamo tako za povezani, če sta bližji kot 1 meter, oziroma za slepi ulici, če sta oddaljeni za več kot en meter.



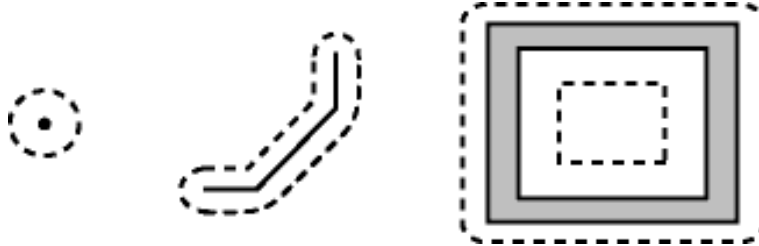
Slika 7.6: Primer uporabe pasu evklidske oddaljenosti za povezovanje cest.

Zgoraj omenjen primer ceste in železnice ponazarja slika 7.7, kjer končnim točkam linij očrtamo krožnice s polmerom 50 metrov (oddaljenost od ceste zaradi parkirišča, nakladalne rampe ipd).



Slika 7.7: Primer uporabe pasu evklidske oddaljenosti za povezovanje ceste in železnice.

Koncept lahko razširimo tudi na poligone in linije. Kot vidimo na sliki 7.8, je rezultat operacije vedno poligon (ali zbirka poligonov) z zaokroženimi robovi. Lahko si predstavljamo, da povezave med točkami namesto skozi točke potegnemo po zunanjih lokih krožnic - po delu, ki leži v zunanosti geometrijskega objekta.



Slika 7.8: Pas evklidske oddaljenost za točko, linijo in poligon [16].

Pas evklidske oddaljenosti se uporablja tudi za implementacijo posebne relacije $a.vRazdalji(b, n)$. Geometrijskemu objektu b dodamo pas oddaljenost velikosti n , da dobimo geometrijski objekt Bb ter testiramo relacijo $a.Seka(Bb)$.

7.5 Topološka usklajenost ali pravilnost prostorskih podatkov

Topološka pravilnost vektorskih podatkov je element kakovosti z najbolj striktnimi omejitvami in se mu pogosto posveča največ pozornosti. Nekatere napake v topologiji geometrijskih objektov lahko vodijo do nepredvidljivih rezultatov in napačnega delovanje programske opreme in analitičnih funkcij. Na primer: barvanje površine nezaprtega poligona lahko vodi do razlitja barve izven meja tega poligona.

Zaradi pomembnosti topološke pravilnosti večina programskih orodij za urejanje prostorskih podatkov že omogoča avtomatsko preverjanje le-te ob vnosu podatkov. Ker podatki pogosto prihajajo iz različnih virov, jih moramo običajno vseeno dodatno preveriti pred uvozom v podatkovno bazo in izdelavo analiz nad njimi. Topološko nepravilne podatke je običajno najboljše zavriniti ali poskusiti avtomatsko odpraviti.

Poleg topoloških napak moramo omeniti še topološke anomalije, ki pogosto nastanejo pri nenatančni digitalizaciji vektorskih podatkov ali pri izdelavi presekov med podatkovnimi nizi. Anomalije običajno ne vplivajo na delovanje programske opreme in so nekatere lahko popolnoma upravičene in predstavljajo dejansko stanje. Vseeno pa so nekakšen pokazatelj topološke usklajenosti ter tudi položajne natančnosti in kot take nezaželene.

Delitev med napake in anomalije ni popolnoma določena in je odvisna od podatkovnega niza ter njegove uporabe, zato jih bomo tudi v naslednjem pod-

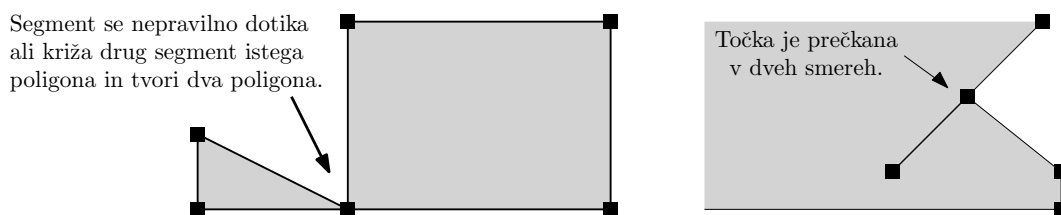
poglavju našteji skupaj. Na strani uporabnika je odločitev, da se na podlagi poročila o topološki usklajenosti odloči, kaj bo klasificiral kot anomalijo ter kaj kot napako in če bo takšne podatke uporabil.

7.5.1 Pogoste topološke napake in anomalije

Povratna zanka ali samo-presek (angl. *Loop backs or self intersections*)

Do te napake pride, ko:

- meja poligona križa samo sebe. Napaki rečemo tudi “metuljast poligon” ali “osmica” (slika 7.9 levo) ali
- linija križa vozlišče dvakrat v različnih smereh (slika 7.9 desno).



Slika 7.9: Primer povratne zanke ali samo-preseka

Napake je potrebno odpraviti z ročnim urejanjem v GIS okolju, saj avtomatični pristop k odpravljanju teh napak ni mogoč, oziroma bi stanje zelo verjetno še poslabšal.

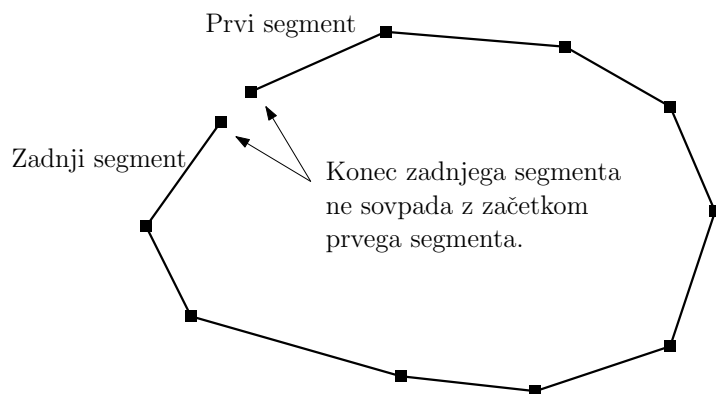
Nezaprt poligon ali obroč

Obroči poligona morajo imeti isto začetno in končno točko - konec zadnjega segmenta in začetek prvega segmenta morata sovpadati. Če to ne drži, pravimo, da je poligon ali obroč nezaprt (slika 7.10).

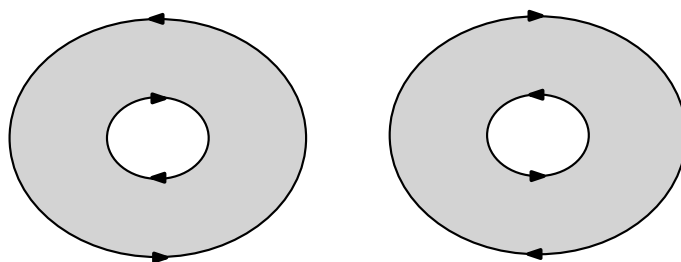
Z avtomatsko proceduro lahko zagotovimo, da bosta prva in zadnja točka identični.

Nepravilna orientacija obročev

Zunanji obroči poligonov morajo imeti točke zapisane, oziroma morajo biti orientirani v smeri urinega kazalca, notranji obroči pa morajo biti orientirani v obratni smeri urinega kazalca (slika 7.11).



Slika 7.10: Primer nezaprtega obroča.



Slika 7.11: Primer pravilne (levo) in napačne (desno) orientacije obročev.

Napako lahko odpravimo z avtomatskim obračanjem orientacije problematičnih obročev.

Presekane linije

Če topologija linij zahteva, da se vse linije stikajo v vozliščih, potem nobena linija ne sme sekati druge na črti.

Napako lahko odpravimo z avtomatsko proceduro, ki na mestu preseka naredi novo vozlišče in razpolovi obe liniji.

Nična geometrija

Kadar zapis objekta nima geometrije, se to lahko šteje kot napaka.

Take zapise lahko avtomatsko odstranimo.

Prazen del geometrije

Do te napake pride pri zbirkah geometrijskih objektov, ko je eden izmed delov v zbirki prazen, kar obravnavamo kot napako.

Prazne dele iz zbirk lahko avtomatsko odstranimo.

Podvojene točke

Test je običajno izveden brez tolerance, zato imajo podvojene točke točno iste koordinate. Pri poligonih in linijah to pomeni sovpadanje dveh zaporednih točk. Pri točkovnih geometrijah pa se takšne točke štejejo za prekrivanja. V kolikor pri testu uporabimo toleranco, rezultat meritve ne bo enak, saj bo odkril tudi tako imenovane "kratke segmente".

Podvojene točke lahko odstranimo z avtomatskimi procedurami, če ne uporabimo tolerance, se meje geometrij ne bodo spremenile.

Kratek segment

Do te napake pride, ko sta dve zaporedni vozlišči na liniji ali obroču poligona bližje od neke predpisane kratke razdalje (npr. 0,05 metra).

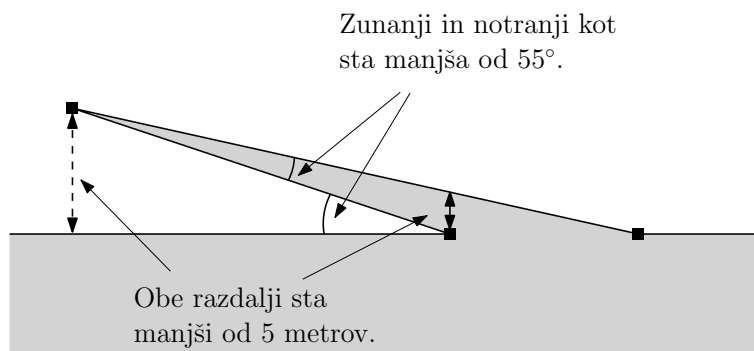
Napako odpravimo z odstranitvijo enega od vozlišč. Lahko uporabimo avtomatsko proceduro, ki ji podamo kriterij za izbiro vozlišča, ki naj bo odstranjen.

Povratna točka (angl. *Kickback*)

Do te napake pride, ko:

- sta notranji in zunanji kot med dvema segmentoma manjša od nekega določenega kota (npr. 55°) in
- sta razdalja med zunanjo špico in zunanjo mejo ter razdalja med notranjo špico in notranjo mejo poligona manjši od neke kratke razdalje (npr. 5 metrov) (slika 7.12).

Napake lahko odstranimo z avtomatsko proceduro, ki odstrani problematično točko in tako odreže špico, vendar je zaradi pogostih lažnih zadetkov bolj priporočljivo ročno popravljanje teh napak.

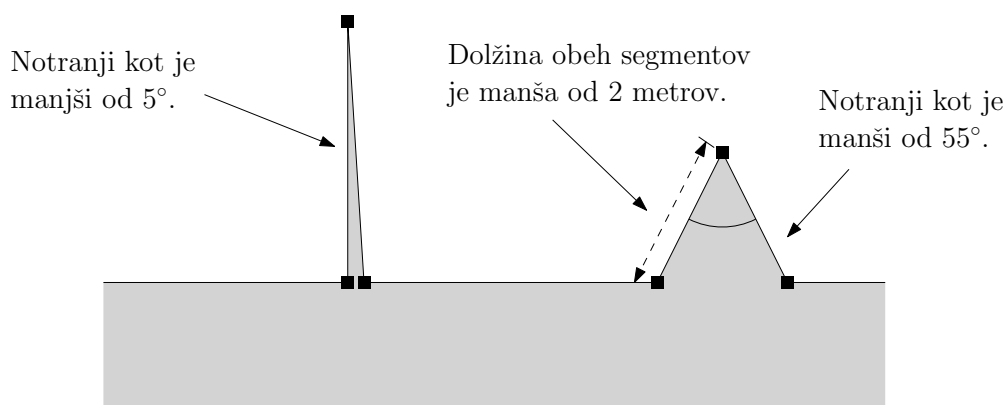


Slika 7.12: Primer povratne točke.

Špica (angl. *Spike*)

Do te napake pride:

- ko je notranji kot med segmentoma, ki tvorita špico, manjši od nekega majhnega kota (npr. 5°) (slika 7.13 levo)
- ali ko je notranji kot med segmentoma manjši od nekega večjega kota (npr. 55°), dolžina segmentov pa je manjša od neke kratke razdalje (npr. 2 metra) (slika 7.13 desno).



Slika 7.13: Primer špice.

Za odpravljanje teh napak velja enako kot za odpravljanje povratnih točk.

Majhna površina

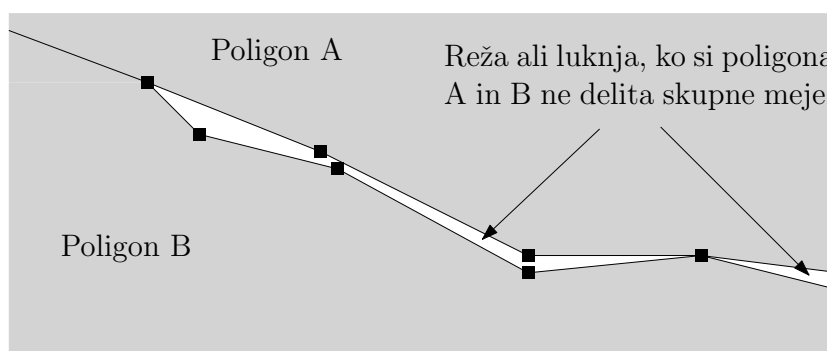
Kot majhne površine se štejejo poligoni, ki imajo površino manjšo od neke vnaprej določene površine (npr. 10 kvadratnih metrov).

Z avtomatsko proceduro lahko te poligone odstranimo ali pa jih pripojimo večjim poligonom.

Reža ali luknja (angl. *Sliver or Gap*)

Do te napake pride:

- ko je površina notranjega obroča (luknje) po združitvi dveh poligonov manjša od neke določene površine (npr. 2 kvadratna metra)
- ali ko je površina luknje numerično manjša kot njen obseg (slika 7.14).



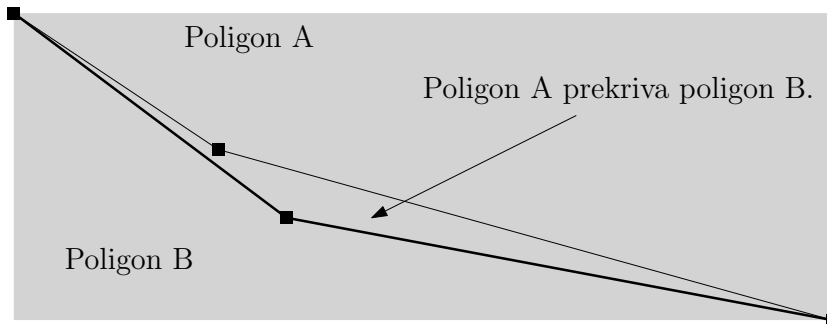
Slika 7.14: Primer reže ali luknje.

Napako se lahko odpravi ročno ali z avtomatsko proceduro, ki po določenih kriterijih spoji sosednje meje.

Prekrivanje poligonov

Če presek dveh poligonov ni prazen, pravimo, da se poligona prekrivata (ne glede na velikost prekrite površine) (slika 7.15).

Napako se lahko odpravi ročno ali z avtomatsko proceduro, ki po določenih kriterijih spoji sosednje meje, če razdalje med njimi padejo v določeno toleranco.



Slika 7.15: Primer prekrivanja poligonov.

Podvojeni pojavi

Pojavi oziroma objekti so podvojeni, ko imajo natanko iste lastnosti - tako geometrijo kot tudi atributne podatke, razlikujejo se le v unikatnem identifikatorju.

Napako odpravimo z avtomatskim izbrisom enega od vnosov, če le-ta ni referenciran preko tujega ključa.

7.5.2 Algoritmi za preverjanje topološke pravilnosti

7.5.2.1 Iskanje presekov med linijskimi segmenti

Linijski segment je konveksna ovojnica dveh točk, poimenovanih končni točki segmenta, ali poenostavljeno: linijski segment sta dve med seboj povezani točki. Imamo množico n segmentov definiranih z x in y koordinatami njihovih končnih točk, zanima nas, kateri od teh segmentov se sekajo.

Najprej pogledjmo primer za $n = 2$, saj je problem seveda trivialen za $n \leq 1$. Če nas ne zanima točka presečišča, lahko preverimo kar ali sta točki prvega segmenta na nasprotnih straneh drugega segmenta. To lahko ugotovimo s preprosto metodo za preverjanje smeri treh točk. Točki a in b segmenta ab ležita na nasprotnih straneh segmenta cd , če in le če natanko en od trojčkov a, c, d in b, c, d leži v nasprotni smeri urinega kazalca. Točko preseka pa lahko izračunamo iz determinant:

$$Px = \frac{\begin{vmatrix} | & x_a & y_a & | & | & x_a & 1 & | & | \\ | & x_b & y_b & | & | & x_b & 1 & | & | \\ | & x_c & y_c & | & | & x_c & 1 & | & | \\ | & x_d & y_d & | & | & x_d & 1 & | & | \end{vmatrix}}{\begin{vmatrix} | & x_a & 1 & | & | & y_a & 1 & | & | \\ | & x_b & 1 & | & | & y_b & 1 & | & | \\ | & x_c & 1 & | & | & y_c & 1 & | & | \\ | & x_d & 1 & | & | & y_d & 1 & | & | \end{vmatrix}} \quad (18)$$

$$Py = \frac{\begin{vmatrix} | & x_a & y_a & | & | & y_a & 1 & | & | \\ | & x_b & y_b & | & | & y_b & 1 & | & | \\ | & x_c & y_c & | & | & y_c & 1 & | & | \\ | & x_d & y_d & | & | & y_d & 1 & | & | \end{vmatrix}}{\begin{vmatrix} | & x_a & 1 & | & | & y_a & 1 & | & | \\ | & x_b & 1 & | & | & y_b & 1 & | & | \\ | & x_c & 1 & | & | & y_c & 1 & | & | \\ | & x_d & 1 & | & | & y_d & 1 & | & | \end{vmatrix}} \quad (19)$$

Ali poenostavljeno:

$$P(x, y) = \left(\frac{(x_a y_b - y_a x_b)(x_c - x_d) - (x_a - x_b)(x_c y_d - y_c x_d)}{(x_a - x_b)(y_c - y_d) - (y_a - y_b)(x_c - x_d)}, \right. \\ \left. \frac{(x_a y_b - y_a x_b)(y_c - y_d) - (y_a - y_b)(x_c y_d - y_c x_d)}{(x_a - x_b)(y_c - y_d) - (y_a - y_b)(x_c - x_d)} \right) \quad (20)$$

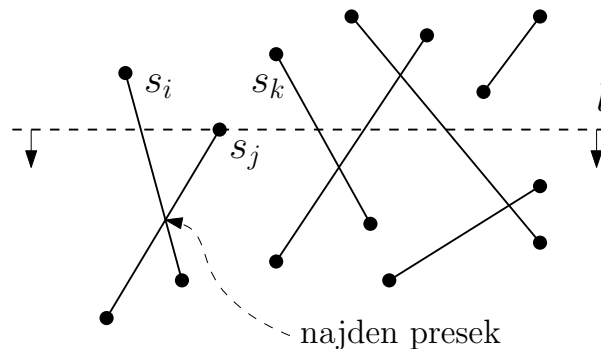
Če želimo za presečišča upoštevati tudi primere, ko se segmenta dotikata v končnih točkah ali ko se končna točka enega segmenta dotika drugega na liniji, moramo dodatno preveriti razdalje med točkami in razdalje med točkami in linijami.

Razširitev problema na $n > 2$ se zdi preprosta, testiramo lahko kar vsak par segmentov in zapišemo segmente, ki se sekajo. Tak algoritem bi bil seveda časovne zahtevnosti $O(n^2)$. Ker preseke obravnavamo, kot napako v topologiji, običajno pričakujemo relativno majhno število presekov, zato potrebujemo algoritem, ki bi se na takšnih podatkih obnašal hitreje, algoritem katerega časovna zahtevnost ne bi bila odvisna le od velikosti vhodnih podatkov (števila linijskih segmentov), temveč tudi od velikosti izhoda (število presekov).

Da bi se izognili testiranju vseh segmentov, moramo upoštevati geometrijo problema - kandidati za preseke so le segmenti, ki ležijo dovolj blizu. Definirajmo y -interval segmenta kot njegovo ortogonalno projekcijo na y -os. Ko se y -intervala dveh segmentov ne prekrivata, lahko rečemo, da sta v y smeri dovolj narazen in se tako ne moreta sekati. Tako moramo testirati le segmente, katerih y -intervali se prekrivajo, se pravi segmente za katere obstaja horizontalna premica, ki preseka oba segmenta. Da bi te segmente našli, si

predstavljamo horizontalno “prebirno premico” l , ki jo premikamo po ravnini v vertikalni smeri in evidentiramo segmente, ki jo sekajo ter jih uredimo od leve proti desni. Takšno urejeno zaporedje predstavlja stanje prebirne premice, ki se lahko spremeni le, ko premica naleti na eno od končnih točk segmenta ali na točko preseka med dvema segmentoma, te točke poimenujemo “dogodkovne točke”.

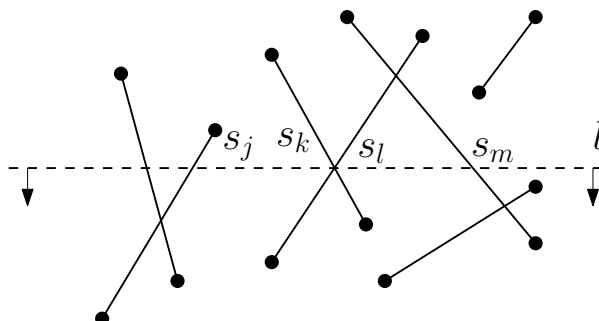
Premico tako pomikamo le med dogodkovnimi točkami. Ko premica naleti na zgornjo končno točko segmenta, ta segment začne sekati premico, zato ga dodamo v stanje premice in testiramo za presek med njegovim levim in desnim sosedom. Če se segmenta sekata pod premico, evidentiramo presek in točko preseka označimo kot novo dogodkovno točko. Npr., če imamo na premici sosednja segmenta s_i in s_k in premica doseže zgornjo točko segmenta s_j , moramo s_j testirati za presek s s_i in s_k (Slika 7.16).



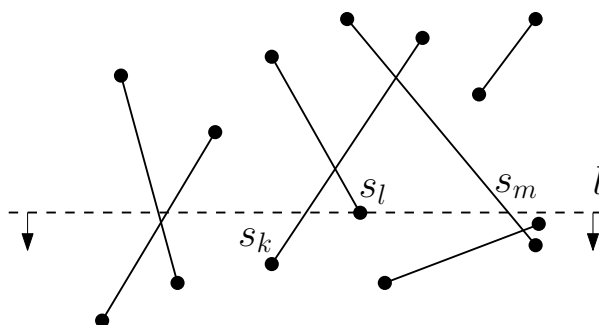
Slika 7.16: Prebirna premica doseže zgornjo končno točko segmenta [1].

Ko premica naleti na dogodkovno točko preseka med dvema segmentoma, zamenjamo vrstni red teh dveh segmentov. Vsak od teh segmentov lahko dobi po največ enega novega soseda, ki jih tudi testiramo za presek. Npr., če imamo na premici štiri segmente s_j , s_k , s_l in s_m , ko ta doseže dogodkovno točko preseka med s_k in s_l , potem s_k in s_l zamenjata položaj ter za presek testiramo tudi med pari segmentov s_l , s_j in s_k , s_m (Slika 7.17).

Ko premica naleti na spodnjo končno točko segmenta, ga odstranimo iz stanja premice in za presek testiramo novo dobljena soseda. Npr., če imamo na premici tri segmente s_k , s_l in s_m , ko ta doseže spodnjo končno točko s_l , bosta segmenta s_k in s_m postala sosednja, zato jih testiramo za presek (Slika 7.18).



Slika 7.17: Prebirna premica doseže presek med segmentoma [1].



Slika 7.18: Prebirna premica doseže spodnjo končno točko segmenta [1].

Algoritem se konča, ko premica naleti na najnižjo končno točko in iz stanja tako odstrani zadnji segment.

Za prvotno urejanje segmentov po njihovi y -smeri potrebujemo $O(n \log n)$ časa. Če za hranjenje stanja prebirne premice uporabimo primerno podatkovno strukturo, na primer uravnoreženo binarno drevo, lahko operacije dodajanja, odstranjevanja in iskanja sosedov dosežemo v času $O(n \log n)$. Za m dogodkovnih točk dobimo $O(m \log n)$, oziroma $O((n+k) \log n)$, kjer je n število končnih točk in k število dobljenih presekov. Časovna zahtevnost algoritma je torej $O(n \log n + k \log n)$.

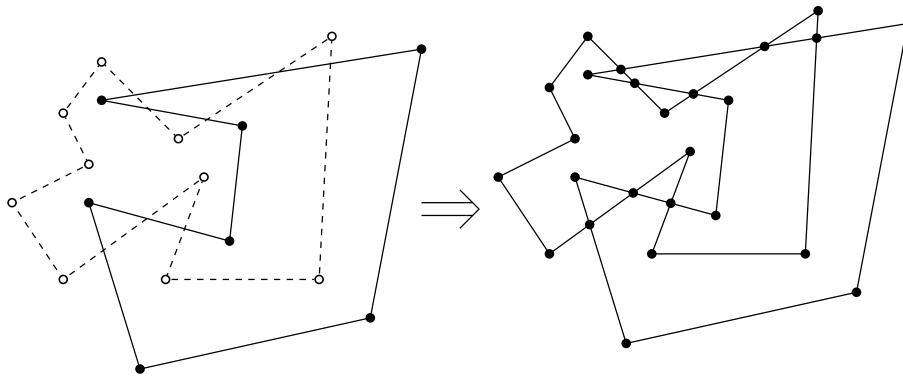
7.5.2.2 Iskanje presekov med geometrijskimi objekti

Ko iščemo preseke med geometrijskimi objekti, nas ne zanimajo le preseki med njihovimi segmenti, temveč rezultat preseka v obliki množice geometrijskih objektov. Imamo množico n geometrijskih objektov S , zanima nas množica geometrijskih objektov P , ki vsebuje preseke med temi geometrijami.

Tako zaradi poenostavitve kot tudi povprečnega hitrejšega delovanja algoritma, bomo zbirke geometrijskih objektov razbili na posamezne geometrijske objekte in namesto zbirk v S dodali te.

Najprej pogledjmo primer za $n = 1$, zanimajo nas torej samo-preseki geometrij. V primeru točke je problem trivialen, saj točka ne more sekati same sebe. Pri linijah in poligonih pa povezane točke obravnavamo kot segmente, preseke pa poiščemo z algoritmom prebirne premice, omenjenim v prejšnjem poglavju. Tudi pri $n > 1$ vsako geometrijo najprej preverimo za samo-preseke. Po definiciji poligona namreč samo-presek poligona pomeni neveljavno geometrijo. Take poligone zato izločimo iz nadaljnje obravnave.

Pri $n = 2$ najprej preverimo presek med ovojnica geometrij. Ovojnica geometrije je minimalni očrtani pravokotnik geometrije in je definirana s spodnjo levo in zgornjo desno točko, oziroma najmanjšima x in y koordinatama ter največjima x in y koordinatama geometrije. Če se ovojnici geometrij ne prekrivata, se tudi geometriji ne moreta sekati. V primeru točk sta ovojnici kar točki sami in se sekata le, če imata identične koordinate. Za iskanje preseka med linijami in poligoni uporabimo predelan algoritem prebirne premice, ki bo zgradil planarni graf preseka D (Slika 7.19).

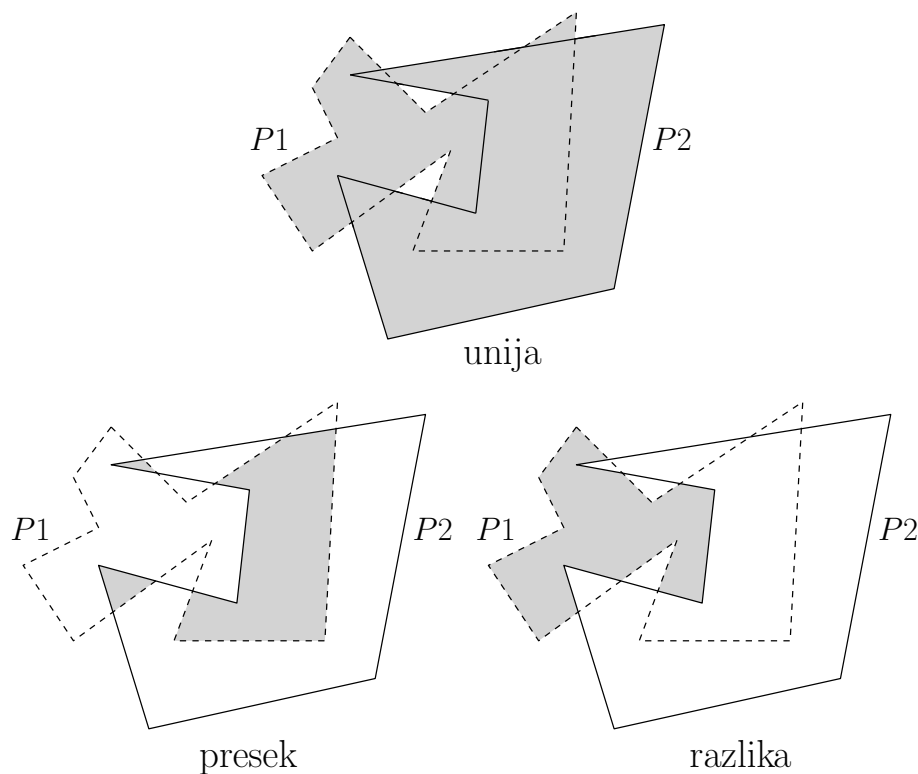


Slika 7.19: Gradnja planarnega grafa preseka.

Povezane točke obeh geometrij obravnavamo kot segmente in jim označimo geometrijo, ki ji pripadajo - $P1$ ali $P2$. Ko bo prebirna premica predelanega algoritma naletela na presek med segmentoma, bomo segmenta razpolovili v točki preseka - nastali bodo štirje novi segmenti, če se segmenta sekata na črti, oziroma dva nova segmenta, če se eden od segmentov dotika končne točke

drugega segmenta (kot presek moramo obravnavati tudi primere, ko se segmenta dotikata na črti). Segmente nad premico dodamo v D , stare segmente (ali en segment v primeru dotikanja na črti) zamenjamo z novimi segmenti pod premico in jim zamenjamo vrstni red. Ko premica naleti na spodnjo končno točko segmenta, ga odstranimo iz stanja premice in dodamo v D .

V dobljenem planarnem grafu D poiščemo množico segmentov, ki so hkrati v $P1$ in $P2$. Če je dobljena množica prazna, preseka ni. V nasprotnem primeru iz dobljenih segmentov sestavimo geometrijo, ki je dobljen presek oziroma prekrivanje med geometrijama. Rezultat lahko uporabimo za preverjanje prostorskih relacij. Iz grafa D pa lahko dobimo tudi unijo $P1 \cup P2$ in razliko $P1 \setminus P2$ dveh geometrij (Slika 7.20).



Slika 7.20: Unija, presek in razlika dveh poligonov $P1$ in $P2$.

Ko je dobljeni presek $P1 \cap P2$ enak $P1$ in $P2$, smo našli podvojeno geometrijo. Če imamo opravka z zbirkami geometrij, moramo seveda preveriti še ostale dele zbirke. Če se geometrijska objekta popolnoma prekrivata in imata identične

atribute, govorimo o podvojenem pojavu.

Za $n > 2$ kot zunanjo zanko ponovno uporabimo princip prebirne premice, vendar namesto segmentov obravnavamo geometrijske objekte oziroma njihove ovojnice. Ko premica, ki se pomika po y -osi, naleti na zgornjo točko ovojnice, jo primerjamo z vsemi ovojnici, ki trenutno sekajo premico - potrebno je primerjati le x -koordinate. Če se dve ovojnici prekrivata, poiščemo presek med geometrijama z uporabo algoritma za $n = 2$.

Ker ovojnic, ki sekajo prebirno premico, nimamo urejenih od leve proti desni, je časovna zahtevnost zunanje zanke $O(n^2)$, vendar je operacija iskanja presekov ovojnic trivialna v primerjavi z operacijo iskanja preseka med dvema geometrijama. Zato zahtevnost zunanje ovojnice zapišimo kot $O(m)$, kjer je m število ovojnic, ki se prekrivajo - v najslabšem primeru, ko se vse ovojnice prekrivajo, je $m = n^2$. Zahtevnost iskanja preseka med dvema geometrijama je $O(k \log k + o \log k)$, kjer je k število segmentov oziroma točk v obeh geometrijah, o pa število presekov med segmenti, poimenovano tudi kompleksnost preseka. Sestavljanje geometrije iz planarnega grafa lahko dosežemo v času $O(k \log k + o \log k)$ [1]. Časovna zahtevnost celotnega algoritma je tako $O(m(k \log k + o \log k))$.

Problem, ko je m blizu n^2 in se delovni čas algoritma približa zgornji asimptotični meji, pa žal ni le teoretičen. Včasih imamo opravka s kompleksnimi geometrijami, ki se raztezajo po veliki površini, njihove ovojnice pa se pogosto prekrivajo (npr. povodja, batimetrija, geološki podatki, ...). Deloma smo to že izboljšali, ker namesto geometrijskih zbirk delamo s posameznimi geometrijami. Lahko pa gremo še korak naprej in velike geometrije razrežemo na manjše dele. V praksi se dobro izkaže metoda bisekcije, kjer geometrije režemo na pol, enkrat po x in drugič po y -smeri, vse dokler posamezne geometrije nimajo manj kot n točk - v zadnjem poglavju omenjeno orodje *cSVT* na primer uporablja privzeto vrednost $n = 2000$. Pri uporabi te metode na koncu ne smemo pozabiti zlepiti presekov istih geometrij.

7.5.2.3 Iskanje lukenj in rež

Pri digitalizaciji ali pri lepljenju poligonov pogosto pride do topoloških anomalij, ko sta si meji poligonov, ki bi se morali stikati, zelo blizu in med njima nastane ozek in dolg prazen prostor, ki mu rečemo reža (angl. *sliver*). Ko imamo opravka s podatki, ki s poligoni predstavljajo pokritost Zemljinega površja

na določenem območju, pa želimo, da v pokritosti ni lukenj in je tako vsak kos površine klasificiran (primeri takšnih podatkov so digitalni kataster, raba zemlje, geologija ipd.).

S stališča algoritma gre pri zgoraj omenjenih primerih za isti dogodek. Med n poligoni moramo poiskati vse prazne prostore in jih glede na tolerančne parametre ustrezno klasificirati kot luknje ali reže. Reže imajo običajno majhno površino in majhno razmerje med površino in obsegom.

Problema se običajno lotimo z gradnjo unije vseh poligonov, da dobimo enega ali več večjih poligonov. Luknje oziroma notranji obroči v poligonih unije so iskane luknje ali reže v podatkih.

Pri gradnji unije poligonov P si ponovno pomagamo s prebirno premico, vendar le toliko, da vse poligone uredimo od zgoraj navzdol in od leve proti desni ter enega po enega dodajamo v trenutno unijo. S tem zagotovimo boljši povprečni delovni čas algoritma, saj bi z naključnim izbiranjem poligonov v prostoru dobili kompleksnejše vmesne unije, dodajanje novih poligonov v unijo pa bi bilo tako počasnejše.

Unijo gradimo tako, da za vsak poligon zgradimo nov planarni graf poligona in trenutne unije, iz katerega kot unijo obeh geometrijskih objektov izračunamo novo unijo. Za prvotno urejanje potrebujemo $O(n \log n)$ časa. Kot smo ugotovili v prejšnjem poglavju, za gradnjo grafa in unije dveh geometrij potrebujemo dvakrat $O(k \log k)$ časa, kjer je k število segmentov oziroma točk v trenutni uniji in trenutnem poligonu. Časovna zahtevnost celotnega algoritma je tako $O(n(k \log k))$.

Povprečni delovni čas algoritma bi sicer lahko zmanjšali tako, da bi gradili le en planarni graf vseh poligonov brez računanja vmesne unije. To pa bi pomenilo, da moramo imeti v pomnilniku točke vseh poligonov hkrati, kar pa v realnosti zaradi velike količine običajno ni sprejemljivo.

7.5.2.4 Iskanje špic in povratnih točk

V primerjavi s prejšnjimi algoritmi sta algoritma za iskanje špic in povratnih točk precej bolj preprosta in časovno manj zahtevna. Kot je očitno iz definicij obeh napak (poglavje 7.5.1), do le-teh lahko pride le pri geometrijah, sestavljenih iz povezanih točk, torej pri linijah in poligonih ter njihovih zbirkah. Pri

obeh algoritmih geometrije obravnavamo kot zaporedja povezanih točk, skozi katere se moramo sprehoditi in pregledati okolico vsake točke.

Pri iskanju špic najprej izračunamo kot, ki ga tvorijo trenutna točka in njeni sosednji točki (kot med dvema segmentoma). Če je ta kot manjši od predpisanega spodnjega kota, smo našli špico. Če je kot večji od predpisanega spodnjega kota, a še vedno manjši od predpisanega zgornjega kota, moramo izračunati še dolžini obeh segmentov. Če sta obe dolžini manjši od predpisane, smo našli špico (slika 7.13).

Ko iščemo povratne točke, moramo primerjati štiri sosednje točke oziroma tri sosednje segmente. Recimo, da primerjamo trenutno točko z njenimi tremi predhodnicami. Tretja točka predstavlja povratno točko, če sta kota med prvim in drugim segmentom ter drugim in tretjim segmentom manjša od predpisanega kota in če sta pravokotni razdalji med prvim segmentom in tretjo točko ter med drugo točko in tretjim segmentom obe manjši od predpisane minimalne razdalje (slika 7.12).

Časovna zahtevnost obeh algoritmov je $O(n)$, kjer je n skupno število točk v vseh geometrijah.

Zaradi podobnosti obeh algoritmov in samih napak obstaja možnost, da isto napako zaznamo v obeh algoritmih. Parametre algoritmov (kote in razdalje) je potrebno nastaviti tako, da najdemo največje število pravih napak in čim manj lažnih zadetkov.

7.6 Absolutna položajna natančnost

Ocena absolutne položajne natančnosti nam podaja stopnjo ujemanja lokacije pojavov v podatkih z njihovo stvarno lokacijo v realnem svetu. Za oceno natančnosti potrebujemo referenčne oziroma "kontrolne podatke", ki jih običajno dobimo z dodatnimi meritvami vzorčnih točk na terenu ali iz podatkovnega niza boljše položajne natančnosti. Lokacije kontrolnih podatkov lahko obravnavamo kot stvarne oziroma natančne ali pa pri oceni upoštevamo že znano natančnost kontrolnih podatkov - pri izmeri dodatnih kontrolnih točk nam to podajajo že naprave same.

Problem povezovanja zapisov med ocenjevanimi in kontrolnimi podatki je enak kot pri določanju presežkov in izostankov ter tematske natančnosti. Ko ocen-

jevanje izvajamo le s kontrolnimi točkami, ne glede na geometrijsko obliko pojavov, se povezovanje ustreznih točk običajno izvede ročno.

Vsaki ocenjevani točki (x_1, y_1) določimo absolutno napako tako, da izračunamo njeno oddaljenost d od kontrolne točke (x_2, y_2) :

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \quad (21)$$

Ko smo izračunali absolutno napako vseh kontrolnih točk, lahko z različnimi statističnimi metodami ocenimo položajno natančnost celotnega podatkovnega niza. Ocena natančnosti je najpogosteje podana s korenem srednje kvadratne napake (*RMSE*) in 95% ter 99% intervalom zaupanja.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n ((x_{2,i} - x_{1,i})^2 + (y_{2,i} - y_{1,i})^2)}{n}} \quad (22)$$

Med pogostimi oblikami podajanja ocene natančnosti so tudi najmanjša in največja napaka, povprečna napaka, odstotek točk z višjo napako od predpisane sprejemljive napake ipd.

Ko razpolagamo s kontrolnimi podatki za celotne pojave - z vsemi točkami na liniji ali poligonu - posameznih točk pa ne moremo ustrezno povezati, lahko natančnost ocenimo z *metodo pasu evklidske oddaljenosti (buffer method)* ali s *hausdorfferjevo razdaljo*. Obe metodi ocenjujeta stopnjo skladnosti oblik in ne določate napak v položaju posameznih točk.

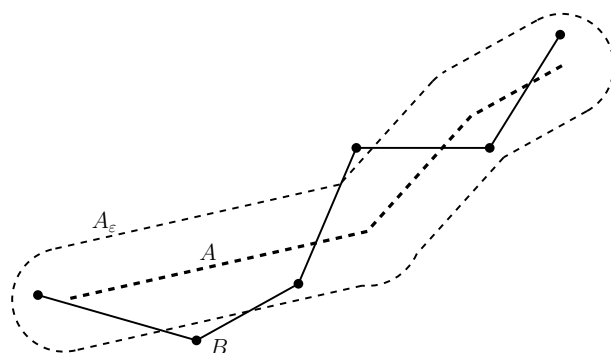
Pri obeh metodah geometrijske objekte najprej razbijemo na posamezne linije. Pri prvi metodi liniji iz kontrolnih podatkov A ("resnični liniji") določimo pas evklidske oddaljenosti širine ε , A_ε (glej poglavje 7.4.3) in preštejemo število ali odstotek točk iz ocenjevane linije B , ki leži znotraj pasu (slika 7.21).

Zgornja metoda ni občutljiva na izstopajoče velike napake, zato je včasih bolj primerna metoda hausdorfferjeve razdalje, ki upošteva celotno dolžino ocenjevanih linij. Če je ε_1 širina pasu evklidske oddaljenosti kontrolne linije A_ε , znotraj katerega padejo vse točke ocenjevanje linije B in ε_2 širina pasu evklidske oddaljenosti ocenjevane linije B_ε , znotraj katerega padejo vse točke kontrolne linije A (slika 7.22), je $\min(\varepsilon_1, \varepsilon_2)$ hausdorfferjeva razdalja d_h :

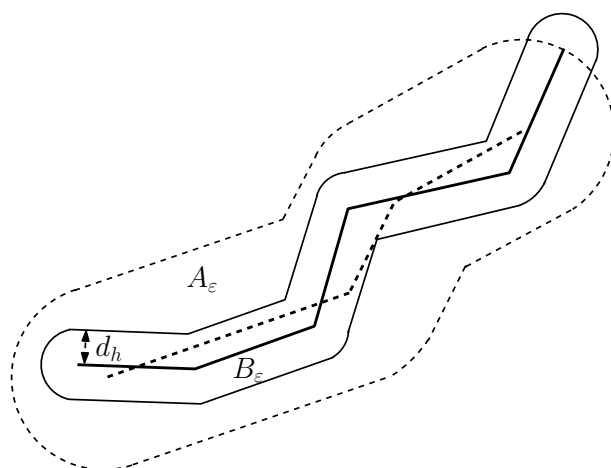
$$d_h(A, B) = \max_{a \in A} \left\{ \min_{b \in B} \{d(a, b)\} \right\}, \quad (23)$$

$$d_h(A, B) \leq \varepsilon \Leftrightarrow A \subset B_\varepsilon \wedge B \subset A_\varepsilon. \quad (24)$$

Za določanje položajne napake celotnih pojavov se včasih uporabljajo tudi mere kot so razlika v dolžini linij, razlika obsega in površine poligonov, razlika v položaju težišč poligonov ipd. Položajno natančnost celotnega niza pa ocenimo z enakimi metodami kot z absolutno napako posameznih točk, le da razdaljo med točkami nadomestimo s izbrano mero napake.



Slika 7.21: Ocenjevanje absolutne natančnosti linije z metodo pasu evklidske oddaljenosti.



Slika 7.22: Ocenjevanje absolutne natančnosti linije s hausdorfferjevo razdaljo.

8 Ocenjevanje kakovosti digitalnega katastrskega načrta Slovenije

V zadnjem delu si bomo z dovoljenjem Geodetske uprave Republike Slovenije (GURS) ogledali primer ocenjevanja kakovosti prostorskih podatkov na podatkovnem nizu digitalnega katastrskega načrta Slovenije (DKN).

Oblika katastrskih parcel v DKN-ju je predstavljena s poligoni, ki so lahko nadaljnje razdrobljeni na poligone parcelnih delov. Vsaka parcela pripada eni katastrski občini (KO), ki je lahko razdeljena v dele KO. Zaradi načina vzdrževanja DKN, ki se izvaja znotraj ene KO, lahko pričakujemo logično neusklajenost parcel na mejah KO-jev. Tak način vzdrževanja je posledica inicialnega zajema katastra iz različnih vrst osnovnih katastrskih načrtov, ki so bili izdelani v različnih merilih in z različnimi vrstami izmere. Na območjih izvajanja komasacijskih postopkov (združitvev zemljišč vseh posestnikov na določenem področju v enoten kompleks) se parcele v celotnem katastru lahko prekrivajo, saj se v času izvajanja teh postopkov znotraj KO vodita dva sloja. Ker je v tem času veljaven del parcel iz enega in del parcel iz drugega sloja, morata biti v celotnem DKN-ju vklopljena oba sloja.

Ocenjevanje je bilo izvedeno z orodjem cSVT (*client Spatial Validation Tool*), izdelanim za potrebe vladnega oddelka za okolje, hrano in podeželje v Veliki Britaniji (*DEFRA - Department for Environment, Food and Rural Affairs*). Orodje implementira večino metod, opisanih v poglavju 7. Orodje sem v času obveznega praktičnega izobraževanja na visokošolskem študiju razvil v podjetju *Sinergise d.o.o.*

Ocenjevali smo le formatno, domensko in topološko usklajenost, saj z referenčnimi podatki nismo razpolagali in smo bili tako omejeni le na izvajanje notranjih metod ocenjevanja kakovosti. DKN smo ocenjevali kot samostojen podatkovni niz, zato pomenska usklajenost v tem primeru nima pomena.

Za ocenjevanje topološke usklajenosti smo uporabili naslednje tolerančne parametre:

- Povratna točka - kot: 55°
- Povratna točka - razdalja: 1 m
- Špica - spodnji kot: 5°
- Špica - zgornji kot: 55°

- Špica - razdalja: 5 m
- Majhna površina: 1 m²
- Kratek segment: 0.05 m
- Zgornja površina reže: 5 m²
- Podvojene točke: 0.004 m

Po ocenjevanju kakovosti na vseh 6.830.178 parcelnih delih smo dobili naslednje rezultate:

- Podatkovni niz je formatno in domensko popolnoma usklajen.
- V podatkovnem nizu je bilo najdenih 194.666 možnih topoloških napak ali anomalij. V tabeli 8.1 je povzeto število posameznih topoloških napak in anomalij. Razporejenost napak po prostoru in nekaj primerov napak je prikazanih na slikah v dodatku A.

Tabela 8.1: Število najdenih topoloških napak in anomalij v DKN.

Topološka napaka ali anomalija	Št. pojavitev
Povratna zanka	0
Nezaprt poligon ali obroč	0
Nepravilna orientacija obročev	0
Presekane linije	0
Nična geometrija	20
Prazen del geometrije	0
Podvojene točke	0
Kratek segment	17.632
Povratna točka	895
Špica	104.992
Majhna površina	3.462
Reža	11.481
Luknja	136
Prekrivanje poligonov	50.850
Podvojen pojav	5.198

Vidimo lahko, da v podatkovnem nizu ni hudih topoloških napak, kot so povratna zanka, nezaprt poligon, nepravilna orientacija obročev ali presekane linije.

Pri podrobnem ogledu anomalij špic se izkaže, da so te v veliki večini primerov upravičene in predstavljajo dejansko stanje - predvsem gre za parcele vzdolž cest. Za boljši rezultat bi bilo potrebno spremeniti tolerančne parametre za določanje špic in ponovno izvesti ocenjevanje, kar pa zaradi dolgotrajnosti postopka ni bilo možno. Če iz skupnega števila napak in anomalij izločimo vse špice, je vseh napak in anomalij le 89.674.

Pri podrobnem pregledu anomalij povratnih točk in kratkih segmentov se izkaže, da so vse posledica nenatančne digitalizacije.

Kot je bilo pričakovati, smo našli precejšnje število prekritih poligonov in rež. Vsa prekrivanja so na mejah KO-jev ali na območjih komasacij. Tudi vse reže so na mejah KO-jev. Našli smo nekaj lukenj znotraj KO-jev.

Nepričakovano smo našli precejšnje število podvojenih pojavov z enako geometrijo in atributi. Ti so verjetno posledica napake pri združevanju DKN-jev posameznih KO-jev v celoten DKN.

Ugotovili smo, da je topološka usklajenost znotraj posameznih KO-jev dobra. Pojavljajo se predvsem anomalije, ki so posledica nenatančne digitalizacije parcel. Problematičnih je le nekaj lukenj znotraj KO-jev in parcele brez geometrije (po podrobnem preverjanju vidimo, da parcele brez geometrije niso tiste, ki bi morale "pokriti" luknje znotraj KO-jev).

Topološka usklajenost celotnega DKN-ja se izkaže za zelo slabo. To je glede na način vzdrževanja katastra sicer razumljivo, težava je predvsem v tem, da se tega zaveda redko kateri uporabnik tega podatkovnega niza. GURS v metapodatkih za DKN, objavljenih na spletnem portalu *prostor.gov.si*, okvirno navaja le položajno in tematsko natančnost ter časovno usklajenost. Celotna logična usklajenost je navedena le kot: "Povezava med lokacijsko in pisno bazo je parcelna številka" [15].

Kot smo ugotovili v uvodnih poglavjih, je poznavanje kakovosti prostorskih podatkov kritično za primerno uporabo le-teh. Nepoznavanje kakovosti podatkov ima lahko zelo negativne posledice in vodi v vrsto napačnih odločitev. GURS bi v metapodatkih moral nujno podrobneje navesti kakovostne značilnosti celotnega DKN-ja.

9 Zaključek

V diplomski nalogi smo predstavili koncept kakovosti prostorskih podatkov in izpostavili pomembnost te kakovosti ter posledice, ki jih lahko ima nepoznavanje in nezavedanje kakovosti prostorskih podatkov. Pregled standardov s področja kakovosti prostorskih podatkov je pokazal, da ti standardi ne definirajo nobenih konkretnih metod za ocenjevanje kakovosti, temveč le definirajo elemente kakovosti, predpisujejo splošne smernice in metodologije ocenjevanja kakovosti ter določajo način poročanja o kakovosti.

Pri iskanju literature za opisovanje konkretnih metod avtomatskega ocenjevanja kakovosti se je izkazalo, da razen metod za ocenjevanje prostorske natančnosti le-te niso nikjer opisane na način, ki bi omogočil dosledno implementacijo. Zato smo pri opisovanju metod uporabili večinoma lastne izkušnje, predvsem izkušnje, dobljene z razvojem orodja cSVT. Opisane metode so le nekatere izmed možnih, vendar zadostujejo za oceno kakovosti večine vektorskih podatkovnih nizov. Za ocenjevanje nekaterih prostorskih podatkov so zaradi drugačne narave podatkov potrebne specializirane metode. Pri opisu metod smo se omejili le na vektorske podatke, saj je pristop k ocenjevanju kakovosti rastrskih podatkov povsem drugačen in zelo odvisen od načina zajema podatkov.

Ocenjevanje in poročanje o kakovosti prostorskih podatkov je žal bolj izjema kot stalna praksa. Najverjetneje je ravno to razlog za pomanjkljivo zavedanje uporabnikov o pomembnosti kakovosti prostorskih podatkov: ker uporabniki ne vedo, da potrebujejo informacije o kakovosti, jih od proizvajalcev in ponudnikov podatkov tudi ne zahtevajo. Ker povpraševanje spodbuja ponudbo, se začaran krog tako sklene.

Priloge

A Slike topoloških napak in anomalij najdenih pri ocenjevanju DKN

V tej prilogi se nahajajo slike, ki grafično prikazujejo rezultate ocenjevanja podatkovnega niza DKN v poglavju 8. Slike od A.1 do A.7 prikazujejo prostorsko razporejenost posameznih vrst napak in anomalij na celnem območju Slovenije. Iz Slike A.8 lahko vidimo, da so prekrivanja in reže v veliki večini razporejene po mejah KO-jev. Na sliki A.9 je prikazana ena izmed najdenih lukenj znotraj KO-ja. Slika A.10 prikazuje primer ene izmed najdenih povratnih točk.



Slika A.1: Prekrivanja v DKN na območju cele Slovenije.
(približno merilo: 1:2.500.000)



Slika A.2: Luknje v DKN na območju cele Slovenije.
(približno merilo: 1:2.500.000)



Slika A.3: Reže v DKN na območju cele Slovenije.
(približno merilo: 1:2.500.000)



Slika A.4: Podvojeni pojavi v DKN na območju cele Slovenije.
(približno merilo: 1:2.500.000)



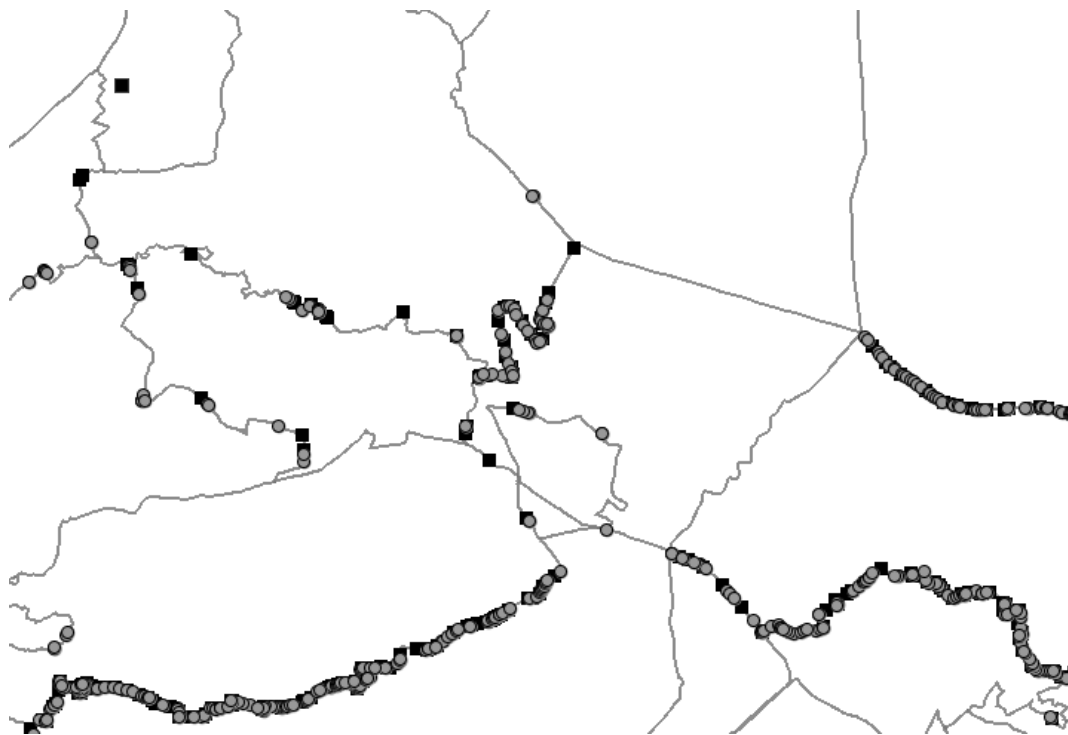
Slika A.5: Povratne točke v DKN na območju cele Slovenije.
(približno merilo: 1:2.500.000)



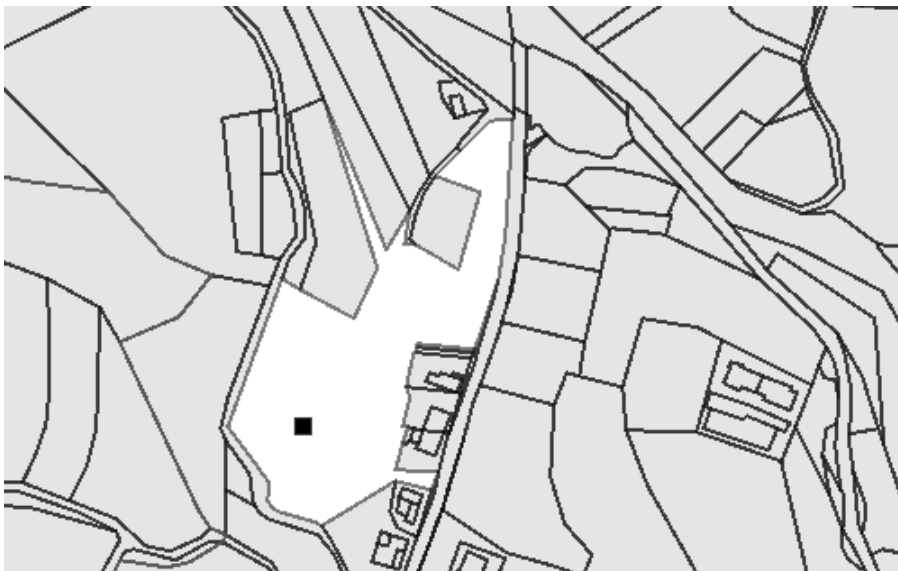
Slika A.6: Kratki segmenti v DKN na območju cele Slovenije.
(približno merilo: 1:2.500.000)



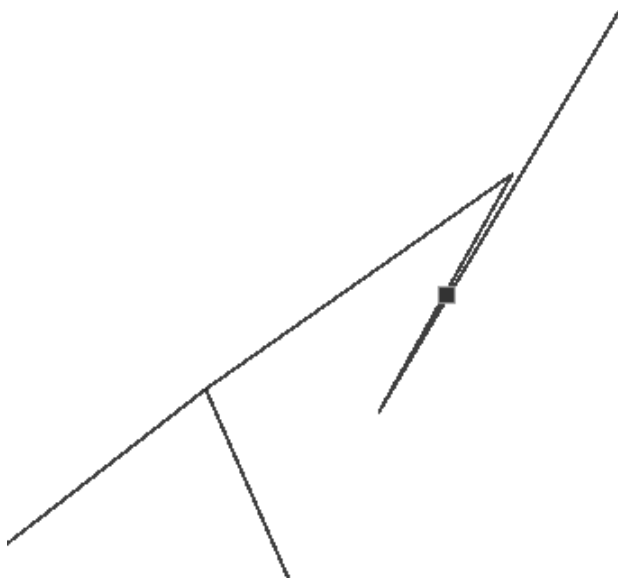
Slika A.7: Majhne površine v DKN na območju cele Slovenije.
(približno merilo: 1:2.500.000)



Slika A.8: Razporejenost rež in prekrivanj po mejah KO-jev.
(približno merilo: 1:50.000)



Slika A.9: Primer luknje znotraj KO.
(približno merilo: 1:4.000)



Slika A.10: Primer povratne točke v DKN.
(približno merilo: 1:50)

Slike

3.1	Prekrivanje cest iz podatkovnih nizov v različnih merilih [2].	7
4.1	Poenostavljen koncept interne in eksterne kakovosti.	9
7.1	Primer pomenske neusklajenosti med cesto in stavbo.	21
7.2	Koncept notranjosti, meje in zunanosti geometrijskega objekta.	22
7.3	Razširitev <i>tangentno</i>	28
7.4	Razširitev <i>meji na</i>	29
7.5	Razširitev <i>striktno</i>	30
7.6	Primer uporabe pasu evklidske oddaljenosti za povezovanje cest.	31
7.7	Primer uporabe pasu evklidske oddaljenosti za povezovanje ceste in železnice.	31
7.8	Pas evklidske oddaljenost za točko, linijo in poligon [16].	32
7.9	Primer povratne zanke ali samo-preseka	33
7.10	Primer nezaprtega obroča.	34
7.11	Primer pravilne (levo) in napačne (desno) orientacije obročev.	34
7.12	Primer povratne točke.	36
7.13	Primer špice.	36
7.14	Primer reže ali luknje.	37
7.15	Primer prekrivanja poligonov.	38
7.16	Prebirna premica doseže zgornjo končno točko segmenta [1].	40
7.17	Prebirna premica doseže presek med segmentoma [1].	41
7.18	Prebirna premica doseže spodnjo končno točko segmenta [1].	41
7.19	Gradnja planarnega grafa preseka.	42
7.20	Unija, presek in razlika dveh poligonov $P1$ in $P2$	43
7.21	Ocenjevanje absolutne natančnosti linije z metodo pasu evk- lidske oddaljenosti.	48
7.22	Ocenjevanje absolutne natančnosti linije s hausdorfferjevo raz- daljo.	48
A.1	Prekrivanja v DKN na območju cele Slovenije.	53
A.2	Luknje v DKN na območju cele Slovenije.	54
A.3	Reže v DKN na območju cele Slovenije.	54
A.4	Podvojeni pojavi v DKN na območju cele Slovenije.	55
A.5	Povratne točke v DKN na območju cele Slovenije.	55
A.6	Kratki segmenti v DKN na območju cele Slovenije.	56
A.7	Majhne površine v DKN na območju cele Slovenije.	56
A.8	Razporejenost rež in prekrivanj po mejah KO-jev.	57
A.9	Primer luknje znotraj KO.	58
A.10	Primer povratne točke v DKN.	58

Tabele

7.1	Osnovna oblika modela DE-9IM [10].	23
8.1	Število najdenih topoloških napak in anomalij v DKN.	50

Literatura

- [1] M. de Berg, O. Cheong, M. van Kreveld, M. Overmars, *Computational Geometry, Algorithms and Applications*, 3. izdaja, Berlin: Springer-Verlag, 2008, pogl. 2.
- [2] R. Devillers, R. Jeansoulin, *Fundamentals of Spatial Data Quality*, London: iSTE, 2006.
- [3] G. J. Hunter, "What's wrong with data quality information?", v zborniku *Proc. GIScience 2000*, Savannah, 2000, str. 201-203.
- [4] ISO 9000:1987, *Quality Management and Quality Assurance - Guidelines for Selection and Use*, International Organization for Standardization (ISO), 1987.
- [5] ISO 8402, *Quality Management and Quality Assurance - Vocabulary*, International Organization for Standardization (ISO), 1994.
- [6] ISO/TC 211, 19113, *Geographic Information - Quality Principles*, International Organization for Standardization (ISO), 2002.
- [7] ISO/TC 211, 19114, *Geographic Information - Quality evaluation procedures*, International Organization for Standardization (ISO), 2003.
- [8] ISO/TC 211, 19115, *Geographic Information - Metadata*, International Organization for Standardization (ISO), 2003.
- [9] ISO/TC 211, 19139, *Geographic information - Metadata - XML schema implementation*, International Organization for Standardization (ISO), 2007.
- [10] OpenGIS *Implementation Specification for Geographic information - Simple feature access - Part 1: Common architecture*, verzija 1.2.0, Open Geospatial Consortium (OGC), 2006.
- [11] W. Shi, *Principles of Modeling Uncertainties in Spatial Data and Spatial Analyses*, Boca Raton: CRC Press, 2008, pogl. 15.
- [12] Direktiva 2007/2/ES Evropskega parlamenta in Sveta o vzpostavitvi infrastrukture za prostorske informacije v Evropski skupnosti (INSPIRE), UL L, št. 108, 2007.

- [13] Uredba komisije (ES) št. 1205/2008 o *izvajanju Direktive 2007/2/ES Evropskega parlamenta in Sveta glede metapodatkov*, UL L, št. 326, 2008.
- [14] (2010) INSPIRE Annex I data specifications testing, Call for Participation. Dostopno na: http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/INSP_testing-CfP_final.pdf
- [15] (2010) Metapodatki digitalnega katastrskega načrta Slovenije. Dostopno na: http://prostor.gov.si/cepp/GURS_izpisiso.jsp?ID=%7B1614DDB9-5216-11D2-BC1C-00A0C9067C11%7D
- [16] (2010) Oracle Spatial Developer's Guide, Spatial Concepts. Dostopno na: http://download.oracle.com/docs/cd/B28359_01/appdev.111/b28400/sdo_intro.htm