

KAKOVOST PROSTORSKIH PODATKOV

Matjaž Ivačič

MOP-Zavod RS za prostorsko planiranje, Ljubljana

Prispelo za objavo: 10.2.1994

Izvleček

V članku je opisana zveza med prostorskimi podatki, ki jih uporabljamo v geografskih informacijskih sistemih (GIS) in njihovo kakovostjo. Uporabniki prostorskih podatkov morajo upoštevati omejeno kakovost prostorskih podatkov pri aplikacijah. Podani so osnovni parametri in obrazložitev kakovosti prostorskih podatkov. Predstavljen je tudi način za izbiro primernih metod ugotavljanja kakovosti prostorskih podatkov in izvedenih rezultatov.

Gljučne besede: GIS, kakovost, prostorski podatki

Abstract

The paper describes the relation among spatial data used in Geographic Information Systems (GIS) and their quality. Users of spatial data must take into consideration the limited quality of spatial data in applications. Basic parameters of spatial data quality are described and explained in short. Some methods for spatial data quality determination are presented.

Keywords: GIS, quality, spatial data

KAKOVOST IN PROSTORSKI PODATKI

V članku želim predstaviti nekaj osnovnih pogledov na kakovost prostorskih podatkov. Namenjen je vsem, ki že uporabljajo prostorske podatke, bodisi v CAD ali GIS okolju, in so večkrat nezadovoljni z njimi, saj ne izpolnjujejo njihovih pričakovanj. Tudi za prostorske podatke velja eno od osnovnih pravil digitalnih podatkov, ki se v žargonskem prevodu glasi: „smeti noter, smeti ven”. Svet, ki nas obdaja, je poln najrazličnejših prostorskih pojavov. Z našim znanjem smo sposobni na najmanjšem delu zemeljske površine zajeti neskončno mnogo prostorskih informacij. Z delom teh informacij se ukvarja tudi geodezija, ki s prostorskimi podatki, ki jih nudi uporabnikom, predstavlja nekakšen most med realnim svetom in modelom realnega sveta, ki je prikazan na klasični karti ali v modernejši podatkovni bazi. Model prostora je generalizirana, okrnjena podoba realnosti, v katerem so predstavljeni realni ali abstraktni elementi, kot na primer gozdovi, reke, posesti, onesnažene površine, območja urbanih con itd. Večine od teh ne moremo eksaktno določiti, ampak so posledica človekovega dojemanja pojavov, ki nas obkrožajo.

Načrti in karte so bili dolga stoletja edini natančnejši in širšemu krogu uporabnikov dostopni nosilci prostorskih informacij, katerih vsebina je namenjena vidnemu zaznavanju. Zato se kartografija praviloma prilagaja

spodobnostim uporabnikov in jim omogoča pravilno razumevanje sporočila karte. Računalnik „bere“ karto na popolnoma drugačen način. V eni sami podatkovni bazi imamo lahko shranjeno veliko količino prostorskih podatkov, ki jih obdelujemo na najrazličnejše načine. Obdelava prostorskih podatkov dobiva z GIS-i nove razsežnosti, ki jih analogni mediji ne morejo nuditi.

Digitalni prostorski podatki skrivajo v sebi drobno past, ki se je številni uporabniki GIS-ov pri analizah, modeliranju in kartografskih prikazih še ne zavedajo dovolj. Kakovost klasičnih medijev je uporabnikom več ali manj dobro znana. Podana je z merilom, stopnjo generalizacije, metodo zajemanja podatkov, z dimenzijsko stabilnostjo papirja ali folije itd. Uporabniki so večinoma vse te elemente pri svojem delu upoštevali. S prehodom na digitalno obdelavo pa so se zgoraj naštetih parametri kakovosti prostorskih podatkov spremenili. Ker so digitalni podatki večinoma pridobljeni iz kartografskega materiala, je njihova kakovost kvečjemu enaka ali največkrat slabša, kot je v izvirnem materialu. Srečamo se torej s problemom kakovosti prostorskih podatkov v geografskih informacijskih sistemih.

Osnovne zahteve kakovosti industrijskih izdelkov so določene z mednarodnimi standardi (ISO 9000 – 9004), ki jih mora upoštevati vsako uspešno podjetje. Geodezija se že srečuje z marketinškimi razmerami in njenem ugled bo odvisen tudi od kakovosti njenih izdelkov. Nezadostna kakovost prostorskih podatkov pomeni razcep med „proizvajalci“ in „uporabniki“. Slednji so, v opravičilo nam geodetom, tudi premalo poučeni in uporabljajo prostorske podatke v aplikacijah, ki po zahtevnosti presegajo dopustni prag kakovosti prostorskih podatkov. Prepletanje in istočasna navzkrižna uporaba digitalnih podatkov, ki so bili zajeti iz meril 1:5 000, 1:50 000 in celo 1:250 000, v praksi sploh ni redko.

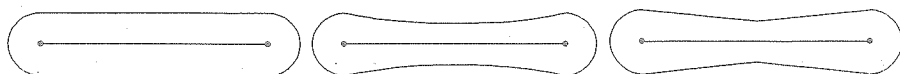
PARAMETRI KAKOVOSTI PROSTORSKIH PODATKOV

Vsi prostorski podatki so le relativno zanesljivi, kar izhaja iz same narave prostorskih pojavov in dinamike sprememb v prostoru. Najprej moramo določiti (identificirati) objekt, ga izmeriti, predstaviti v podatkovnem modelu (vektor, raster), transformirati v izbran referenčni koordinatni sistem, zapisati v digitalno obliko, shraniti v podatkovno bazo itd. Z vsakim od naštetih korakov se spreminja tudi kakovost prostorskih podatkov. Definicijo kakovosti prostorskih podatkov, kot jo opisujejo različni avtorji (Hunter, Fain, SDTS), lahko strnemo v en sam stavek: Kakovostni prostorski podatki so tisti, ki popolnoma zadovoljijo potrebe uporabnika. V svetu je najbolj razširjena naslednja klasifikacija parametrov kakovosti prostorskih podatkov:

- pozicijska natančnost (positional accuracy)
- opisna ali atributna natančnost (attribute accuracy)
- logična združljivost ali skladnost (logical consistency)
- celovitost (completeness)
- zgodovina ali opis minulih dogajanj (lineage).

Geodezija se večinoma ukvarja s pozicijsko natančnostjo. Kakovost podatkov lahko določimo z natančnostjo meritve ali z medsebojnim odstopanjem ponavljajočih se meritev. Pri tem običajno uporabimo normalno (Gaussovo) porazdelitveno funkcijo in natančnost meritev izrazimo s standardno deviacijo. Za

terenska, fotogrametrična in satelitska merjenja poznamo več načinov določevanja pozicijske natančnosti. Klasična geodezija uporablja elipso pogreškov kot grafično ponazoritev pozicijske natančnosti merjene točke. V GIS-ih pa nas pogosteje kakor natančnost točke v prostoru zanima natančnost linije, ki razmejuje različne prostorske pojave. Linijo grafično lahko ponazorimo z epsilon pasom. Epsilon pas je območje, znotraj katerega se nahaja „prava“ linija, in se pogosto uporablja pri določevanju pozicijske natančnosti. To metodo lahko uporabimo tudi za določevanje natančnosti ročne ali polavtomatske digitalizacije.

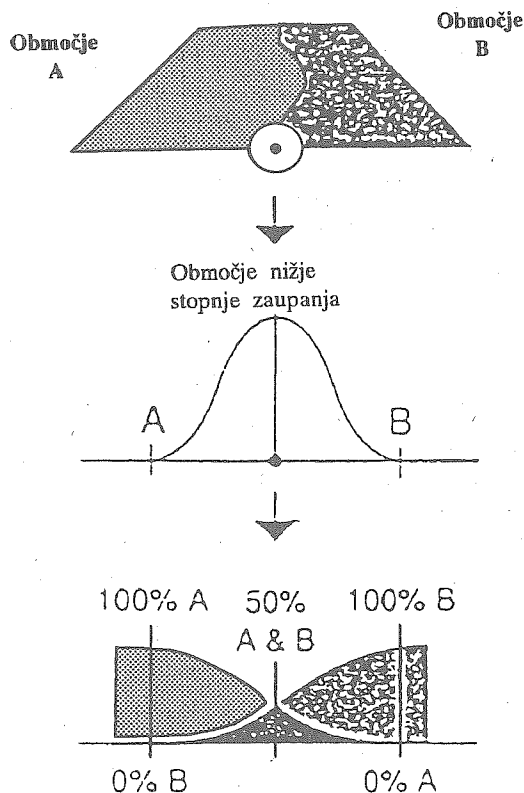


Slika 1: Primeri različnih oblik epsilon pasov

Opisna (atributna) natančnost podatkovnih baz je mogoče zanemarjen parameter, čeprav morda še bolj pomemben. Posamezna prostorska entiteta ima poleg osnovnega identifikatorja in lokacije v prostoru običajno še več opisnih atributov. Ti imajo lahko pomembnejšo vlogo kot zelo natančne koordinate. Atributni podatki se običajno tudi spreminjajo hitreje kot lokacija entitete, zato so z vidika kakovosti lahko bolj pomembni. Atributni podatki so različnih tipov, zato jih moramo pred ugotavljanjem njihove kakovosti natančneje opredeliti. To najlažje storimo, če odgovorimo na naslednja vprašanja (Laurini et al. 1992): ima atribut specifično vrednost, je uvrščen v razred, je atributni podatek originalen ali izveden, je nujno potreben za identifikacijo entitete. Logična združljivost ali skladnost je kakovostni parameter, ki ga vizualno dokaj hitro opazimo. Pojavi se običajno, kadar spajamo informacijske sloje, ki so bili zajeti ločeno, ali kadar ob povečevanju presežemo spodnjo mejo pozicijske natančnosti. V teh primerih opazimo, da nam reke tečejo pod cestami, kjer ni mostov, stanovanjski bloki so sredi križišč, centri stavb padejo zunaj stavbe itd. Takšne probleme moramo reševati predvsem z uporabo najprimernejše metode, ki upošteva razlog nastanka takšnega neskladja. Premikanje posameznih prostorskih entitet brez poznavanja razlogov za različno pozicijo pogosto tvori nova nesoglasja in je kot tako običajno neprimerno.

Celovitost je parameter, ki se običajno navezuje na celotno podatkovno bazo in ne samo na posamezne entitete. Povežemo ga lahko z izrazom ažurnost. Aronoff (Aronoff 1989) razlikuje tri tipe celovitosti podatkovnih baz:

- Celovitost informacijskega sloja – samo v idealnem primeru je zastopanost pozicijskih in atributnih podatkov v bazi 100%. Odvisna pa je tudi od dinamike pojava. Geološki podatki so trajnejši od vremenskih.
- Celovitost klasifikacije – v primerih različnih podatkovnih virov pride do različne klasifikacije pojavov. Pojav iz ene podatkovne baze lahko pripada dvem razredom ali nobenemu v drugi bazi in obratno.
- Celovitost preverjanja podatkov – nekateri podatki so merjeni direktno na terenu, drugi so pridobljeni posredno. Primer je interpolacija rezultatov pedoloških meritev.



Slika 2: Atributna natančnost na meji območij A in B (Goodchild 1989)

Nikoli ne moremo zagotoviti 100-odstotne celovitosti podatkovne baze, ampak se temu idealu približujemo. Paziti pa je treba, kadar zahtevajo naše aplikacije ažurnejšo podatkovno bazo, kot nam je na voljo. Kako so bili podatki pridobljeni in kaj se je z njimi dogajalo do časa njihove uporabe, nam pove zgodovina ali opis minulih dogajanj, ki je posredno zelo pomemben parameter kakovosti prostorskih podatkov. Nanaša se na vire podatkov, kdo in kje je zajemal podatke, način digitalizacije (ročna digitalizacija, skeniranje), resolucija skeniranja, uporabljene transformacijske metode, parametri transformacije itd. Običajno se vsi ti podatki kar nekako izgubijo. Pri zajemu iz kartografskih podlag moramo vedeti tudi, kako so bile karte narejene, kakšna je stopnje generalizacije itd. Opis minulih dogajanj je najpomembnejši posredni parameter kakovosti prostorskih podatkov, iz katerega lahko razberemo vzroke za nizko ali visoko kakovost.

UPOŠTEVANJE KAKOVOSTNIH PARAMETROV PRI DELU Z GIS-I

Uporabljač geografske informacijske sisteme kot primerno orodje za podporo prostorskim odločitvam, moramo upoštevati kriterije minimalne kakovosti prostorskih podatkov. Le-ti se spreminjajo od aplikacije do aplikacije, zato jih največkrat tudi ne moremo dovolj eksaktno določiti. Metode ugotavljanja kakovosti

prostorskih podatkov lahko razvrstimo na: metode ugotavljanja kakovosti izvornih prostorskih podatkov in metode ugotavljanja kakovosti izvedenih rezultatov. Pri izvornih podatkih definiramo kakovost točkovnih, linijskih in ploskovnih elementov predvsem z metodami prostorske statistike. Z definiranjem elementov elips pogreškov, širinami in oblikami epsilon pasov lahko določimo pozicijsko natančnost prostorskih entitet. Pri atributni natančnosti se največkrat uporablja statistična verjetnost. Ostale parametre največkrat lahko določimo izkustveno ali na osnovi opravljenih eksperimentov. Ugotavljanje kakovosti izvedenih rezultatov lahko opravimo analitično samo pri določenih operacijah, kot na primer pri prekrivanju dveh slojev. Pri kompleksnih aplikacijah, ki dodobra izkoristijo sposobnosti orodij GIS-ov, tega ne moremo storiti. Zato velikokrat uporabljamo simulacijske metode, ki kljub svoji relativni enostavnosti dajejo dobre rezultate. Znana je simulacijska metoda Monte Carlo, kjer s spreminjanjem vhodnih podatkov analiziramo spremembe na izhodu in tako določimo „občutljivost“ aplikacije. V primeru dovolj kakovostnih podatkov in „neobčutljive“ aplikacije se nam ob spremembi vhodnih podatkov za N% spremeni za N% tudi rezultat. Običajno pa je ta sprememba nelinearna in kakovost pada v odvisnosti od števila in zahtevnosti uporabljenih GIS operacij.

Kakovost prostorskih podatkov postaja z naraščajočim številom GIS-ovih aplikacij vse pomembnejša. Zavedati se moramo, da orodja GIS niso sama sebi namen in da so v glavnem namenjena tehnološki podpori pri prostorskih in drugih odločitvah. Kakovost prostorskih odločitev, ki imajo dolgoročen vpliv na naše okolje, je v veliki meri odvisna tudi od kakovosti vhodnih podatkov in od načina, kako jih uporabljamo.

Viri:

- Aronoff, S., 1989, *Geographic Information Systems: A Management Perspective*, WDL, Ottawa, Ontario, Kanada.
- Fain, M.A., 1991, *Quality: Through Conversion Process Management*, GIS/LIS 1991 Proceedings.
- Hunter, G.J., Beard K., 1992, *Understanding Error in Spatial Databases*, *The Australian Surveyor*, Vol. 37, No. 2.
- Goodchild, M., Gopal, S., 1989, *Accuracy of Spatial Databases*, Taylor & Francis, London.
- Ivančič, M., 1993, *GIS Applications for Physical Planning in Slovenia: The Problem of Spatial Data Quality*, TEMPUS Join European Project Education in Land Information Systems, Tehnična univerza Delft, Nizozemska.
- Laurini, R., in Thompson D., 1992, *Fundamentals of Spatial Information Systems*, Academic Press Limited, London, San Diego.
- SDTS, 1993, *Spatial Data Transfer Standard, Spatial Data Quality*, ZDA.

Recenzija: mag. Božena Lipej
Joc Triglav