



UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI INGEGNERIA

Corso di Laurea Triennale in Ingegneria Informatica

Tesi di Laurea

Controlli per la qualità dei database topografici a grande scala

Relatore: Prof. M. Rumor

Laureando: Alan Stocco

Matricola: 580502 – IF

Anno Accademico 2010-2011

Sommario

L'argomento di questa tesi è lo studio della qualità nei database topografici a grande scala. La definizione generale di qualità secondo la norma ISO 9000 è: "Qualità: Grado in cui un insieme di caratteristiche intrinseche soddisfano i requisiti." La tesi si sviluppa nell'ambito della cartografia e precisamente nella cartografia digitale. L'introduzione degli strumenti tecnologici come i computer nella cartografia hanno portato a importanti cambiamenti nell'approccio al processo cartografico, come l'utilizzo dei database topografici. I database topografici sono dei contenitori organizzati di dati spaziali e sono parte integrante dei Sistemi Informativi Territoriali (GIS). L'architettura di un database geografico si distingue principalmente da quella dei database tradizionali per la capacità di gestire il dato geografico. Il dato geografico è composto da due componenti: la parte geometrica che descrive la forma e la posizione degli oggetti nello spazio e la parte semantica cioè gli attributi alfanumerici caratterizzanti gli elementi riprodotti. Esiste anche il bisogno d'informazioni aggiuntive che diano indicazioni sulle relazioni spaziali tra gli oggetti. Nei database geografici si possono ricavare relazioni topologiche tra gli oggetti come ad esempio: connessione, adiacenza e inclusione.

Perché i database topografici siano degli strumenti utili per l'analisi e la rappresentazione dei dati geografici devono rispettare delle regole. Queste regole sono esplicitate in una serie di specifiche che descrivono le modalità di acquisizione e di strutturazione di tutte le informazioni all'interno di un GIS. Esse definiscono i requisiti che il database deve rispettare riguardanti sia l'aspetto semantico sia l'aspetto geometrico dell'informazione contenuta in esso e impongono vincoli sulla completezza, la conformità al modello dati del contenuto semantico, sull'accuratezza e la precisione del dato geometrico. Impongono anche relazioni e vincoli topologici tra le classi del modello.

Le specifiche definiscono esplicitamente i requisiti di un database: un database che rispetti tutti i vincoli imposti è da considerarsi formalmente corretto. Per verificare se i dati sono corretti si deve definire il concetto di correttezza e come si può verificare. Gli errori sono violazioni delle specifiche in particolare nei database topografici riguardano vincoli topologici e sulle geometrie e per rilevarli esistono degli strumenti appositi.

Esistono errori che non sono rilevabili da questi strumenti. In particolare la tesi poggia sulla constatazione che anche database formalmente corretti presentano degli errori. L'obiettivo di questa tesi è individuare un insieme di questi errori, organizzarli in una tassonomia e fornire eventualmente delle linee guida per la loro individuazione. L'individuazione di questi errori si basa sul concetto di anomalia: un'anomalia è una situazione in cui l'informazione rappresentata in un database cartografico appare, secondo logica comune, errata.

Un'anomalia, pur non corrispondendo ad alcuna violazione dei vincoli formali del database, può effettivamente individuare un errore nel dato. Questo è dovuto a due cause che sono le due facce di una stessa medaglia: da una parte le specifiche falliscono nel fornire una formalizzazione completa dei requisiti necessari a garantire la reale correttezza dell'informazione; dall'altra ci sono degli errori che non sono rilevabili verificando i vincoli imposti.

Il lavoro è stato svolto all'interno del laboratorio GIRTS che dal 2006 si occupa di attività legate ai GIS soprattutto con il progetto di ricerca CARGEN., il cui obiettivo è sviluppare un processo di generalizzazione automatico per la produzione di un database topografico in scala 1:25000 e 1:50000, partendo dal database territoriale in scala 1:5000, fornito dalla regione Veneto.

Questa tesi si sviluppa nei seguenti cinque capitoli:

- **Capitolo 1: Introduzione alla cartografia**
In questo capitolo verranno, per prima cosa, fornite alcune definizioni riguardante i concetti di cartografia, carta geografica e scala; seguirà un breve resoconto sulla storia della cartografia. Successivamente, saranno illustrate in modo generale le varie classificazioni e tipologie delle carte geografiche e l'evoluzione digitale cioè i GIS.
- **Capitolo 2: Analisi della cartografia e della qualità di una cartografia**
Cos'è la qualità di una cartografia e come si valuta. Panoramica sui vincoli espliciti nella formazione dei database, sugli errori sul contenuto semantico, sul modello, sulle geometrie e in generale sulle specifiche.

Definizione di errore topologico e presentazione degli strumenti esistenti per trovarli.

- **Capitolo 3: Analisi della cartografia e della sua qualità**
In questo capitolo sono chiariti gli obiettivi della tesi e l'approccio usato per conseguirli. Sarà fornita la definizione di anomalia, una lista di anomalie e la procedura per creare una classificazione.
- **Capitolo 4: Anomalie**
.Processo di classificazione delle anomalie, tassonomia risultante con esempi.
- **Capitolo 5: Conclusioni**
Questo capitolo chiude la tesi, traendo le conclusioni del lavoro svolto, fornendo alcune implementazioni e idee per trovare delle anomalie e presentando alcuni possibili sviluppi futuri.

Sommario

Capitolo 1 Introduzione.....	7
1.1 Cartografia.....	7
1.2 Evoluzione della cartografia: dalla carta al digitale	9
1.3 Cartografia in Italia	11
1.3.1 IGM	12
1.3.2 CTRN	12
1.3.3 Gis	13
1.3.4 Database Territoriale (DBT)	13
1.3.5 Laboratori GIRTS e progetto CARGEN.....	15
Capitolo 2 Analisi della cartografia e della sua qualità.....	16
2.1 Qualità del modello cartografico.....	17
2.2 Panoramica sulle specifiche di contenuto	18
2.3 Parametri di qualità del modello DBT	18
2.4 Strumenti rilevazione errori topologici	20
Capitolo 3 Aspetti analizzati	22
3.1 Approccio	22
3.2 Obiettivi.....	22
3.3 Lista di anomalie	23
3.3 Tassonomia.....	23
Capitolo 4 Anomalie	28
4.1 Grafi	28
4.1.1. Interruzioni	28

4.1.2 Tratto isolato o troppo corto.....	31
4.1.3 Mancanza vertice comune incrocio.....	31
4.1.4 Incoerenza di tratti adiacenti	32
4.2 Forma	32
4.2.1 Limiti di dimensione legati al valore semantico	33
4.2.2 Forme artificiali troppo irregolari	34
4.2.3 Forme naturali troppo regolari	34
4.2.4 Cappi	37
4.2.5 Fiumi non monotoni.....	37
4.2.5 In generale relazione tra forme e posizione altimetrica	37
4.3 Distribuzione	38
4.3.1 Distribuzioni non uniformi.....	38
4.4 Posizione	38
4.4.1 Posizionamento logico	38
4.4.2 Divergenza punto	38
Capitolo 5 Conclusioni.....	40
5.1 Algoritmi per la rilevazione	40
5.1.1 Interruzioni nei grafi.....	40
5.1.2 Dimensione rispetto al contenuto semantico.....	41
5.2 Conclusioni e possibili sviluppi	42

Capitolo 1 Introduzione

1.1 Cartografia

La cartografia è lo studio e la pratica di creare mappe. Utilizzando le conoscenze scientifiche, tecniche e artistiche si modella la realtà per comunicare informazioni spaziali (vicinanza, posizione, dimensione..ecc). Una mappa è una rappresentazione simbolica della realtà che ne rappresenta ed enfatizza solo alcuni gli aspetti. Le carte geografiche sono degli strumenti molto impiegati nelle più svariate azioni della nostra vita quotidiana. La carta geografica, che quindi è il prodotto di questa riproduzione cartografica, è una rappresentazione grafica riprodotta secondo simboli convenzionali, di una parte o di tutta la superficie terrestre, o di alcuni suoi aspetti particolari.

Due aspetti importanti nella definizione di una carta sono:

- **Destinazione finale.** Possono essere suddivise in carte topografiche e carte tematiche. Le prime sono fatte per scopi generici e rappresentano elementi naturali e artificiali, le seconde rappresentano informazioni specifiche come le precipitazioni o la temperatura.
- **Scala.** Definisce la dimensione dello spazio di rappresentazione ed influenza la quantità, il tipo e la simbologia degli oggetti sulla carta.

Uno degli aspetti cruciali nella cartografia è proprio il concetto di scala. La scala di una mappa è definita come il rapporto tra la distanza sulla mappa e la corrispondente distanza reale. È indicato come 1:n, ove n è il numero di volte per cui risulta ridotta sulla carta la distanza topografica tra due punti. Quindi la scala secondo cui la carta geografica rappresenta la realtà è il rapporto fra le distanze risultanti sulla scala stessa e le corrispondenti distanze reali sulla superficie della Terra, essa determina il grado di dettaglio della carta e la sua precisione.

Così, le carte su grande scala riproducono una piccola zona e possono contenere numerosissimi dettagli; al contrario, le carte su piccola scala coprono un'area più vasta evidenziando un ristretto e selezionato numero d'informazioni.

Le carte si possono quindi suddividere in base ad essa nel seguente modo:

- carte geografiche ($s < 1:1\,000\,000$),
- carte coreografiche ($1:1\,000\,000 < s < 1:200\,000$),
- carte topografiche in senso stretto ($1:200\,000 < s < 1:5\,000$),
- mappe ($1:5\,000 < s < 1:500$)
- piante ($s > 1:500$)



Figura 1: esempio di una stessa zona a differente scala.

1.2 Evoluzione della cartografia: dalla carta al digitale

Dalla cartografia tradizionale realizzata a mano su supporto cartaceo, negli ultimi decenni si è passati, attraverso diverse fasi evolutive, alla cartografia digitale, cioè basata su supporti informatizzati. L'utilizzo dell'informatica ha determinato una progressiva evoluzione della cartografia digitale. Questo processo si può dividere in tre fasi principali:

- Automazione cartografica

Una delle prime applicazioni dell'informatica nel campo della cartografia ha visto l'utilizzo di *software* di *editor* grafici per la fase di compilazione/disegno della carta. In pratica, ci si è serviti di strumenti informatizzati per produrre la cartografia a tratto su carta nella forma classica. I dati geografici numerici così acquisiti hanno una struttura semplice, finalizzata principalmente alla sola riproduzione grafica automatizzata.

- Cartografia numerica

La dicitura «cartografia numerica» è utilizzata per identificare una raccolta di dati geografici in forma digitale, finalizzata al disegno automatico ed alla gestione con metodologie informatiche, in cui è generalmente presente solamente un file grafico dove le informazioni associate agli oggetti rappresentati (es. larghezza, numero di corsie, fondo delle strade) vengono esplicitate attraverso l'uso degli attributi grafici (livello, colore, spessore, tipo di linea, ecc.). Nella cartografia numerica non sono presenti vincoli topologici.

- Database geografici

L'architettura di un database geografico si distingue da quella dei database classici per il tipo di dato che deve gestire: il dato geografico: gli oggetti spaziali non sono più solo una collezione di segni grafici e simboli, ma sono entità dotate di attributi, inserite in gerarchie relazionali, cui corrisponde non solo una rappresentazione grafica, ma anche un contenuto logico e semantico.

Il dato geografico così inteso è utile per attività di analisi spaziale che ha consentito la costituzione di Sistemi Informativi Geografici, noti con l'acronimo inglese GIS (Geographic Information System), con capacità di elaborazione spaziale oltre che di rappresentazione cartografica.

Grazie all'ausilio di sistemi informatici e dei GIS, negli ultimi trent'anni la situazione della cartografia è profondamente cambiata. Innanzitutto le carte sono ora realizzate sulla base delle informazioni contenute in un database: più precisamente, sono dei sottoprodotti del database. Il computer non è più utilizzato solamente per automatizzare le tecniche descrittive dei cartografi, ma è diventato lo strumento di valutazione della qualità dei dati, l'elemento che fonde i dati tra loro e ne valuta la compatibilità. Un GIS è quindi un sistema computerizzato realizzato per l'acquisizione, memorizzazione, trasformazione e visualizzazione di dati spaziali dal mondo reale. Le sue potenzialità derivano dalla capacità di poter mettere in relazione diverse informazioni che si riferiscono allo stesso contesto spaziale, e di poter fornire una soluzione chiara e centrata sulle specifiche necessità dell'utente circa i rapporti vigenti fra di esse.

Per rappresentare dei dati spaziali in un sistema informatico è necessario collezionare, per ogni singolo elemento, una serie d'informazioni che si possono raggruppare in due tipologie:

- Geometriche: con riferimento alla rappresentazione cartografica dell'oggetto, come la sua forma, dimensione e posizione geografica;
- Informative: attinenti agli attributi (numerici, testuali o di altro tipo) associati all'oggetto.

Si ricavano in più informazioni Topologiche: relative alle relazioni spaziali reciproche fra gli oggetti, come ad esempio:

- Connettività
- Inclusione
- Adiacenza

1.3 Cartografia in Italia

Di seguito sarà fornita una panoramica sulla storia recente e la situazione della cartografia in Italia.

In accordo con una legge del 1960 la produzione cartografica in Italia è stata assegnata a due enti: l'Istituto Geografico Militare (IGM) per le scale medie (1:25000 e minori) e il catasto per le grandi scale (1:10000 e maggiori).

Nel 1977 è avvenuto il trasferimento delle funzioni alle regioni in materia di ambiente e territorio, per cui le regioni possono gestire in modo autonomo la creazione delle carte regionali. Questo ha portato alla produzione di mappe regionali a scala 1:5000 chiamate Carta Tecnica Regionale(CTR). Data l'assenza di un modello comune di riferimento ogni regione sviluppò la propria CTR seguendo le proprie esigenze, portando quindi alla creazione, di fatto, di 20 CTR differenti.

Tutte le mappe di quel periodo erano disegnate a mano. Con l'introduzione dei computer le cose cambiarono. Prima i dati della CTR furono importati nell'equivalente digitale, la CTRN. Poi l'utilizzo di strumenti digitali come il database rivoluzionarono il modo di trattare e di concepire il dato geografico.

Nel 2000 fu sancito il passaggio alla cartografia digitale quando iniziò, per opera dell'IGM, la produzione di una nuova linea di mappe, la serie DB25 che utilizza un database per immagazzinare le informazioni. In seguito anche gli uffici cartografici regionali intrapresero la strada della digitalizzazione cartografica, attraverso la realizzazione dei database territoriali (DBT). Sfruttando i DBT viene a crearsi la possibilità di usare la derivazione per produrre la nuova serie 25DB.

Di seguito sono trattati più nel dettaglio gli argomenti sopra citati soffermandosi soprattutto sugli aspetti di interesse per questo lavoro quindi: la cartografia a grande scala(1:5000 e 1:10000), cioè la CTRN per il passato

recente e il DBT per il prossimo futuro, l'ente a cui è affidata la produzione cartografica(cioè l'IGM e le regioni) e alla fine una panoramica sul modello DB25 e la relazione che sussiste tra questo e il DBT.

1.3.1 IGM

L'Istituto Geografico Militare è stato fondato nel 1872 ed è il creatore della prima carta nazionale, la Nuova Carta Topografica d'Italia con scala 1:100000. Attualmente l'IGM mantiene più tipologie di carte che coprono il territorio nazionale, realizzate in scale diverse:

- Per la scala 1:25000 esistono le serie 25, 25V(la vecchia serie) e 25DB.
- Per la scala 1:50000 le serie 50 e 50L che differiscono per il numero di colori.
- per la scala 1:100000 le serie 100V(la vecchia serie) e 100L.

La produzione delle serie 25 dell'IGM, fino a questo momento, si è appoggiata a rilievi eseguiti in gran parte con metodo aerofotogrammetrico, oppure è avvenuta tramite rilievo numerico o analogico cui viene fatta seguire una fase di rappresentazione grafica eseguita con metodologie automatiche o manuali. Un importante avvenimento è stato l'introduzione del modello 25DB basato sull'utilizzo di database. Questo nuovo modello ha dato via a un processo di ammodernamento anche per la cartografia a grande scala gestita dalle regioni.

1.3.2 CTRN

La Carta Tecnica Regionale Numerica è la versione digitale della CTR. Con questo s'intende che i dati della CTR sono stati semplicemente copiati in un file per renderli disponibili in formato digitale questo senza adeguarli o modificarli. In generale la CTRN è una cartografia generale e metrica, in formato vettoriale. Derivando dalla CTR ne eredita la mancanza di standard comuni tra le varie regioni. La carta, la cui produzione ha nel rilievo fotogrammetrico la principale fonte di dati, è disponibile in due scale di rappresentazione: la scala 1:5.000 e la scala 1:10.000.

Gli oggetti e le informazioni territoriali contenute nella CTRN sono organizzati in Livelli e Codici:

- I Livelli costituiscono una primaria classe di aggregazione degli oggetti
- I Codici che descrivono le caratteristiche particolari dei Livelli

Essa è distribuita in due formati: nel formato DXF (Drawing Exchange File format) o nel formato SHP (Shape File). I dati della CTRN soffrono in generale del fatto di essere state realizzate prevalentemente con tecniche CAD, finalizzate unicamente alla stampa tramite plotter, ed essere solo successivamente state elevate al rango di carte digitali. La struttura dei dati, in particolare, deriva da una rappresentazione priva di ogni forma di controllo di coerenza topologica che inficia la possibilità di usare direttamente i dati della cartografia per analisi.

1.3.3 Gis

Un Gis è un sistema informativo computerizzato che permette l'acquisizione, la registrazione, la gestione, l'analisi e la visualizzazione d'informazioni con contenuto geografico spaziale. I GIS insomma permettono di manipolare oggetti spaziali e di estrarne una grande quantità d'informazioni. In questo contesto le mappe diventano solo uno dei prodotti in output di un GIS. Proprio le mappe da elementi puramente grafici e statici si trasformano in oggetti dinamici che vengono man mano aggiornati e contengono molte più informazioni oltre a quelle visibili su carta. Il singolo dato geografico diventa un'entità dotata di attributi e inserita in una gerarchia relazionale; a questa entità corrisponde, oltre che una rappresentazione grafica, un contenuto logico e semantico.

1.3.4 Database Territoriale (DBT)

L'assenza di uno standard unico per tutte le regioni per le CTR e la novità del DB25 portarono anche le regioni a realizzare un nuovo modello basato sulla cartografia digitale: il database territoriale (DBT).

Il modello del DBT è nato nell'ambito del progetto IntesaGis, dove IntesaGis sta per intesa tra Stato, Regioni ed enti locali sui sistemi informativi territoriali.

Queste specifiche definiscono anzitutto i requisiti di un modello digitale del terreno; prendono poi in esame le diverse tecniche di produzione. L'organizzazione degli oggetti all'interno del DBT è realizzata tramite l'uso di classi. Le classi sono raggruppate in temi e questi ultimi sono a loro volta raggruppati in strati. E', però, da evidenziare che tale raggruppamento non è rappresentato in modo esplicito nel database topografico, la cui struttura comprende solo classi. Gli oggetti del modello possono essere: puntuali, lineari, areali ed è richiesta la copertura totale del territorio in forma topologica, senza né sovrapposizioni né "buchi". Nonostante ciò sono previsti dei casi in cui è consentita la presenza di aree sovrapposte che non partecipano alla copertura di base. In più è definito il comportamento degli oggetti dal punto di vista topologico.

Il modello DBT nel futuro sostituirà la cartografia CTRN e, a differenza di questa, usa una modalità di rappresentazione orientata agli oggetti cioè ogni elemento rappresentato nella mappa è un' oggetto geografico dotato di attributi. Per rappresentare dei dati spaziali in un sistema informatico è necessario collezionare, per ogni singolo elemento, una serie d'informazioni che si possono raggruppare in tre tipologie:

- Geometriche: con riferimento alla rappresentazione cartografica dell'oggetto, come la sua forma, dimensione e posizione geografica;
- Informative: attinenti agli attributi (numerici, testuali o di altro tipo) associati all'oggetto.

Sono altresì importanti le relazioni topologiche che si possono ricavare e che riguardano le relazioni spaziali fra gli oggetti come la sovrapposizione, l'adiacenza, l'inclusione e altre.

Questa fondamentale differenza fra CTRN e DBT fa sì che sfruttando i DBT viene a crearsi la possibilità di usare la derivazione per produrre la nuova serie 25DB.

1.3.5 Laboratori GIRTS e progetto CARGEN

Il lavoro è stato svolto all'interno del laboratorio GIRTS (Geographical Information systems Real-Time Systems) che dal 2006 si occupa di attività legate ai GIS soprattutto con il progetto di ricerca CARGEN (Cartographic Generalization), il cui obiettivo è sviluppare un processo di generalizzazione automatico per la produzione di un database topografico in scala 1:25000 e 1:50000, partendo dal database territoriale in scala 1:5000, fornito dalla regione Veneto.

Capitolo 2 Analisi della cartografia e della sua qualità

Allo scopo di rappresentare e gestire le informazioni spaziali mediante un GIS, è necessario utilizzare una rappresentazione dei dati che sia sganciata dalla realtà fisica; questo è realizzato definendo un modello dei dati che sia abbastanza ampio da accogliere al suo interno tutti gli oggetti che esistono nel mondo fisico e che sia sufficientemente elastico da permettere di adattarlo alla realtà. Rispetto a una rappresentazione puramente geometrica degli oggetti presenti nella realtà, a un GIS è richiesto di gestire, tramite vincoli, tutte le informazioni che riguardano le mutue relazioni spaziali tra i diversi elementi cioè di strutturare i dati definendone anche vincoli topologici. Oltre a questi due aspetti, geometrico e topologico, il modello dei dati deve prevedere l'inserimento al suo interno dei dati descrittivi dei singoli oggetti reali, definibili come attributi. Questi tre insiemi d'informazioni immagazzinati (geometria, attributi) e ricavabili (topologia) vengono poi effettivamente implementati in un GIS mediante uno specifico modello fisico che si basa su strutture dei dati di tipo relazionale, tipiche dei database.

La cartografia si può considerare come uno dei risultati dei GIS. Dunque per analizzare la qualità cartografica bisogna prima chiarire cos'è la qualità del modello e la correttezza del dato cartografico. Entrambi sono di difficile definizione perché non completamente oggettivi. Per iniziare è utile definire la qualità in generale, la definizione data dalla norma ISO 9000 del 2005 è la seguente: "Qualità: Grado in cui un insieme di caratteristiche intrinseche soddisfano i requisiti.". Per essere utile, al pari di ogni grandezza, la qualità deve poter essere misurata. La misura della qualità consiste nel valutare quanto un prodotto è lontano da quello ideale: per farlo occorre quindi considerare le caratteristiche richieste e costruire un metodo che permetta di misurarle.

Di seguito sarà definita la qualità del modello cartografico e quali sono gli indici per misurarla. Sarà poi fatta una panoramica su cosa sono le specifiche di contenuto, le specifiche del modello IntesaGis e i parametri di qualità per il DBT. Per ultimo saranno analizzati gli errori, gli strumenti per rilevarli e se questi strumenti rilevano tutti gli errori.

2.1 Qualità del modello cartografico

Gli indici utilizzati per misurare la qualità di un modello¹ cartografico si basano sulla:

- Accuratezza: misura la corrispondenza tra realtà e modello.
 - Posizionale. Differenza della posizione sulla cartografia rispetto alla sua posizione reale.
 - Descrittiva. Differenza della descrizione rispetto a ciò che rappresenta.
 - Temporale. Differenza tra la data in cui è stato rilevato il dato e quella in cui è utilizzato.
- Precisione: misura la risoluzione del modello.
 - Sulla posizione. Cifre con cui si esprimono le coordinate.
 - Sugli attributi. Finezza di classificazione.
 - Sul tempo. Dettaglio con cui si esprime il tempo.
- Consistenza: rispetto delle regole
 - Spaziale. Conformità alle regole spaziali.
 - Temporale. Conformità alle regole temporali.
 - Tematica. Misura la contraddizione negli attributi tematici.
- Completezza: rappresentazione di tutti gli elementi della realtà d'interesse
 - Spaziale
 - Temporale
 - Tematica

¹ Fonte slide del corso di Sistemi Informativi del prof. M. Rumor

2.2 Panoramica sulle specifiche di contenuto

Le specifiche descrivono le norme di acquisizione e di strutturazione di tutte le informazioni potenzialmente utili all'interno di un sistema informativo territoriale. In particolare le specifiche sulla formazione di un database topografico definiscono i requisiti che il database deve rispettare riguardanti sia l'aspetto semantico sia l'aspetto geometrico dell'informazione. Solitamente descrivono:

- Gli elementi che devono essere rappresentati nel database (classi, attributi, attributi geometrici, associazioni, domini)
 - Le classi di oggetti presenti, da cosa sono composte e che tipo di geometrie le formano.
- I vincoli d'integrità, cioè delle proprietà intrinseche che gli elementi rappresentati devono soddisfare (s'intendono le proprietà verificabili sugli elementi rappresentati stessi, senza osservazioni dirette del mondo reale.)
 - L'elenco degli attributi per ogni classe. Per ogni attributo è definito il tipo di dato e i limiti del valore.
- Elementi descrittivi: qui rientrano tutte le informazioni utilizzabili dagli esseri umani per capire come interpretare un contenuto del database in termini di realtà rappresentata e viceversa come interpretare una situazione reale nei termini degli elementi del database.
- Le relazioni e i vincoli topologici

2.3 Parametri di qualità del modello DBT

Nel capitolo 2.1 sono stati esposti gli indici di qualità per un modello cartografico. Nello specifico la determinazione della qualità del DBT è effettuata mediante l'analisi di alcune grandezze, dette parametri di qualità.

I parametri di qualità presi in esame sono indicati nel documento 1n1014²:

- **Congruenza logica.**
Riguarda la valutazione delle caratteristiche di contenuto del DBT sia per il rispetto dello schema fisico di formato sia per la strutturazione concettuale della componente spaziale.
 - Di formato
 - Di dominio
 - Di geometria
 - Di topologia
- **Accuratezza posizionale**
Si valuta lo scostamento delle coordinate dalla reale posizione sul terreno rispetto alla tolleranza indicata.
- **Completezza**
Fornisce l'attendibilità della sola presenza/assenza di un determinato oggetto topografico nel DBT.
 - Eccesso
 - Omissione
- **Accuratezza tematica**
Valuta la correttezza della codifica degli oggetti del DB e la completa informazione di attribuzione.
 - Di classificazione
 - Toponomastica

La conformità di un insieme di dati alle specifiche indica come e cosa deve esserci nel dato, nel caso dei database, questa conformità è composta di due aspetti:

- **Conformità reale:** la conformità reale riguarda la corrispondenza tra il contenuto informativo e la porzione di mondo reale alla quale il dato si riferisce.
- **Conformità intrinseca:** la conformità intrinseca riguarda la consistenza dell'informazione contenuta nel dato secondo le specifiche.

² http://www.centrointerregionale-gis.it/public/DB_Topografici/1n1014_v_1_4.pdf

2.4 Strumenti rilevazione errori topologici

Forniti gli indici di qualità del modello e le specifiche dei database topografici, resta da verificare la correttezza del dato. Perché un dato sia corretto deve per prima cosa rispettare le specifiche. La valutazione della qualità di un dato deve tener conto sia della conformità reale sia della conformità intrinseca alle specifiche. Nei database topografici i vincoli imposti riguardano principalmente la conformità intrinseca, perché solamente quest'ultima può essere valutata automaticamente.

Esistono degli strumenti per rilevare la violazione di questi vincoli. Sarà preso in considerazione il tool GeoUML Catalogue, realizzato dallo SpatialDBGroup del Politecnico di Milano, con un progetto cofinanziato dalle Regioni, attraverso il CISIS (Centro Interregionale per i Sistemi informatici, geografici e statistici). Il software ideato per la gestione delle specifiche di contenuto dei database topografici contiene il catalogo degli oggetti necessari alla formazione dei DBT e i vincoli. Proprio l'aspetto legato ai vincoli è di maggiore interesse per il tema della tesi. I vincoli in questo strumento sono raggruppati in due tipi:

- “Vincoli Topologici”. Definiscono legami topologici tra gli elementi e quindi legati a questi concetti: adiacenza, sovrapposizione, contenimento, equivale.

Di seguito alcuni esempi di vincoli topologici di GeoUML Catalogue::

- Non devono esistere sovrapposizioni tra nodi idrici.
 - La linea di costa marina deve coincidere con parte di contorno dell'area di mare.
 - Il boundary del tracciato dei canali corrisponde a un insieme di nodi idrici.
 - Non deve esistere sovrapposizione tra i nodi idrici.
- “Vincoli Part whole”. Definiscono vincoli di appartenenza e partizionamento, ad esempio:

- La superficie dell'area stradale è composta di oggetti delle classi area di circolazione veicolare, pedonale, ciclabile ed eventuali manufatti dell'infrastruttura di trasporto.
- Il territorio di una regione è partizionato nel territorio delle province in cui è scomposta.
- Il tracciato della rete tranviaria è costituito dal tracciato degli elementi tranviari, tra loro disgiunti o al più adiacenti; viceversa il tracciato di ogni elemento tranviario deve appartenere alla rete

Esistono tuttavia una serie di errori che non sono rilevabili da questi vincoli e quindi dagli strumenti che li utilizzano.

Capitolo 3 Aspetti analizzati

3.1 Approccio

Nonostante i vincoli esposti nel capitolo precedente, e i tool per rilevarli, nei casi reali è possibile osservare, all'interno di database formalmente corretti, situazioni in cui il dato rappresentato appare errato. L'individuazione di questi errori si basa sul concetto di anomalia: un'anomalia è una situazione in cui l'informazione rappresentata in un database cartografico appare, secondo logica comune, errata.

Un'anomalia, pur non corrispondendo ad alcuna violazione dei vincoli formali del database, può effettivamente individuare un errore nel dato. Questo è dovuto a due cause che sono le due facce di una stessa medaglia: da una parte le specifiche falliscono nel fornire una formalizzazione completa dei requisiti necessari a garantire la reale correttezza dell'informazione; dall'altra ci sono degli errori che non sono rilevabili verificando i vincoli imposti.

L'obiettivo di questa tesi è individuare un insieme di queste anomalie, organizzarle tramite un processo di classificazione in una tassonomia e fornire eventualmente delle linee guida per la loro individuazione. La rilevazione delle anomalie richiede un approccio particolare. L'approccio usato in questa tesi è simile alla progettazione bottom-up. Si parte cioè da dei particolari casi base e poi si crea una classificazione generale, che potrebbe portare all'individuazione di nuovi casi.

3.2 Obiettivi

Dal concetto di anomalia e dall'approccio utilizzato nasce l'obiettivo principale che è creare una procedura per classificare ogni specifica anomalia e creare una tassonomia che raggruppi le anomalie per caratteristiche comuni. Solo in seconda battuta sono fornite delle linee guida per la rilevazione automatica.

3.3 Esempi di anomalie

Prima di organizzare le anomalie si devono trovare. Con questo non s'intende rilevarle sulla mappa ma indicare degli esempi rilevanti che poi portino alla generalizzazione dei casi e a una classificazione ordinata. Per rilevare questi esempi si può far affidamento unicamente alla logica e all'osservazione. Ecco una lista di alcuni esempi di anomalie:

- Zone di mappa prive di edifici.
- Interruzioni nella strada.
- Laghi di forma quadrata.
- Edifici di forma molto spigolosa.
- Curve brusche nei binari ferroviari.
- Strade a una corsia che diventano per brevi tratti autostrade per poi tornare alla feature di partenza.
- Strutture portuali in montagna.
- Fiumi in salita.
- Ponti senza strade vicine.
- Ospedali senza edifici vicini.
- Boschi di piccole dimensioni.
- Muri senza altre strutture vicine ai punti terminali.

3.3 Tassonomia

La tassonomia è la disciplina della classificazione. Con questo termine dunque ci si può riferire alla classificazione gerarchica dei concetti. Nella matematica la classificazione tassonomica prende la forma di una struttura ad albero. A capo della struttura c'è un elemento singolo, cioè il nodo radice, le cui proprietà si

applicano a tutte le altre istanze della categoria, i figli. In questo caso la radice è la caratteristica comune a tutti i casi rilevati cioè essere un'anomalia. I figli di questa radice sono le categorie in cui sono stati divisi gli elementi dell'elenco precedente generalizzandoli e ampliandoli di significato. Per identificare queste categorie si analizza la lista del punto precedente raggruppando i casi in base agli aspetti in comune utilizzando un diagramma di flusso per dividere le anomalie secondo una serie di condizioni binarie. Di seguito le condizioni utilizzate e il diagramma:

1. Se sono locali o globali. Cioè se riguardano l'intera estensione della mappa o una sua parte ristretta.
2. Se trattano un elemento singolo o gruppi di elementi.
3. Se riguardano la distribuzione di tutti gli elementi presenti o solo una parte.
4. Se concernono vincoli sulla posizione o sulla vicinanza con altri elementi.
5. Se riguardano vincoli sulle dimensioni (lunghezza, area) o sulla forma.
6. Se il vincolo sulla dimensione è dovuto alla risoluzione della mappa o alla semantica.
7. Se la forma analizzata è naturale o artificiale.

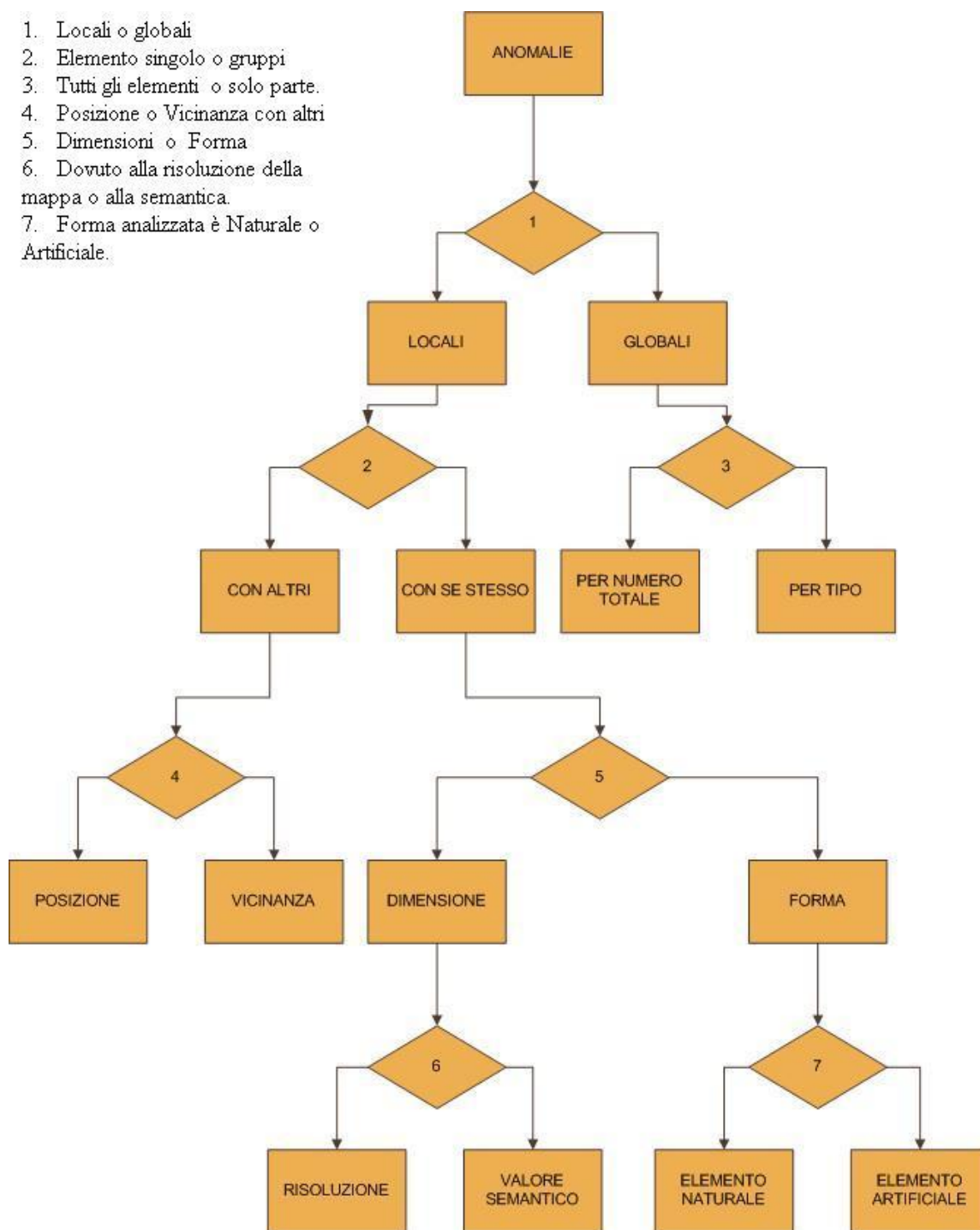


Figura 2 Diagramma di flusso per la classificazione delle anomalie

Il risultato di quest'algoritmo porta a identificare nuove categorie. Di queste alcune sono fuse e ad altre sono tolti dei casi da incorporare in altre categorie. Ne risulta una classificazione in quattro categorie rappresentabili tramite una struttura ad albero. Di seguito un elenco delle categorie definitive e di quali risistemazioni sono frutto:

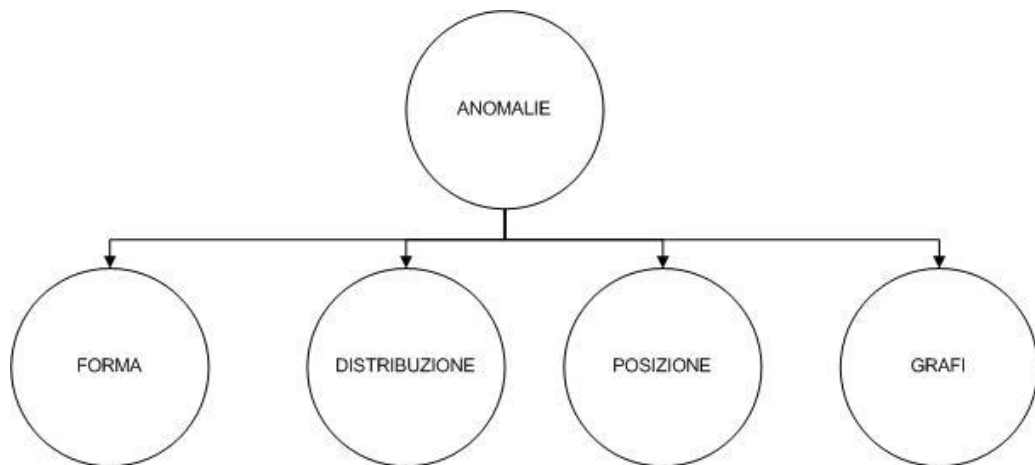


Figura 3 Struttura ad albero rappresentante la classificazione delle anomalie

- **Forma:** questa categoria raggruppa le anomalie riguardanti le forme delle geometrie e le loro dimensioni. Nasce dalla fusione dei casi riguardanti:
 - Forme di elementi naturali troppo regolari. Come ad esempio un lago quadrato.
 - Forme di elementi artificiali troppo irregolari. Come ad esempio una casa troppo spigolosa o un binario che compie percorsi troppo tortuosi.

Ragionamenti basati sulla dimensione essendo questa legata fortemente alla forma. Esistono due tipi di limiti di forma basati sulle dimensioni:

- Limiti legati alla risoluzione. Ad esempio: IGM utilizza 1/10 di millimetro come limite percepibile. Quindi alla scala di 1:25000 0,1mm rappresenta 2,5 metri. Dunque non sono possibili interruzioni inferiori a 2,5 metri nelle strade e laghi di area inferiore a 6,25 metri quadri.
 - Limiti legati al valore semantico. Ad esempio: è accettabile una casa di un metro quadro, perché potrebbe essere una baracca, ma non è pensabile un bosco di due metri quadri.
- **Distribuzione:** contiene le anomalie globali. Analizza la ripartizione degli elementi nella mappa, contiene sia ragionamenti sulla distribuzione totale degli elementi cioè un semplice conteggio del numero, sia analisi sulla distribuzione non uniforme di determinati elementi.
 - **Posizione:** questa categoria fonde le categorie riguardanti la vicinanza e la posizione essendo la vicinanza funzione della posizione nello spazio. La nozione di vicino amplia il concetto topologico tipico che considera solo legami di adiacenza potendo così ampliare il raggio d'azione dell'analisi spaziale. Alcuni casi che fanno parte di questo gruppo sono: strutture portuali in montagna, ponti senza strade vicine o ospedali senza edifici vicini.
 - **Grafi:** in questa categoria sono raggruppate anomalie su strutture a grafo di cui classici esempi sono il grafo stradale o ferroviario. Questa categoria non è frutto del diagramma di flusso ma dalla rilevazione che molte anomalie riguardano questo tipo di struttura a causa delle particolari proprietà, soprattutto esistono molte anomalie riguardanti la connettività.

Capitolo 4 Anomalie

In questo capitolo saranno esplicitate meglio le categorie tassonomiche identificate nel capitolo precedente. Per ognuna sarà fornita una serie di esempi di casi rilevabili.

4.1 Grafi

Questa tipologia riguarda anomalie che inficiano o alterano la connettività dei grafi. Questa categoria è stata creata per raggruppare anomalie di diverso tipo ma tutte relative a strutture a grafo. Questo è stato fatto per sottolineare che molte di queste anomalie sono frutto delle particolari caratteristiche che un grafo deve rispettare come ad esempio la connettività e la coerenza tra i tratti adiacenti. La prima casistica affrontata riguarda le piccole interruzioni nella struttura lineare che forma il grafo, la seconda le dimensioni dei tratti che formano il grafo, la terza gli incroci senza punto in comune e l'ultima incoerenza tra tratti adiacenti.

4.1.1. Interruzioni

Con interruzioni s'intende la presenza di un vuoto di piccole dimensioni tra due elementi lineari che sembrano congiunti, questo porta il fallimento degli algoritmi di "navigazione" della struttura lineare. Un esempio è una strada che s'interrompe bruscamente e riparte dopo una distanza superiore alla risoluzione ma inferiore a un certo valore di soglia fissato.

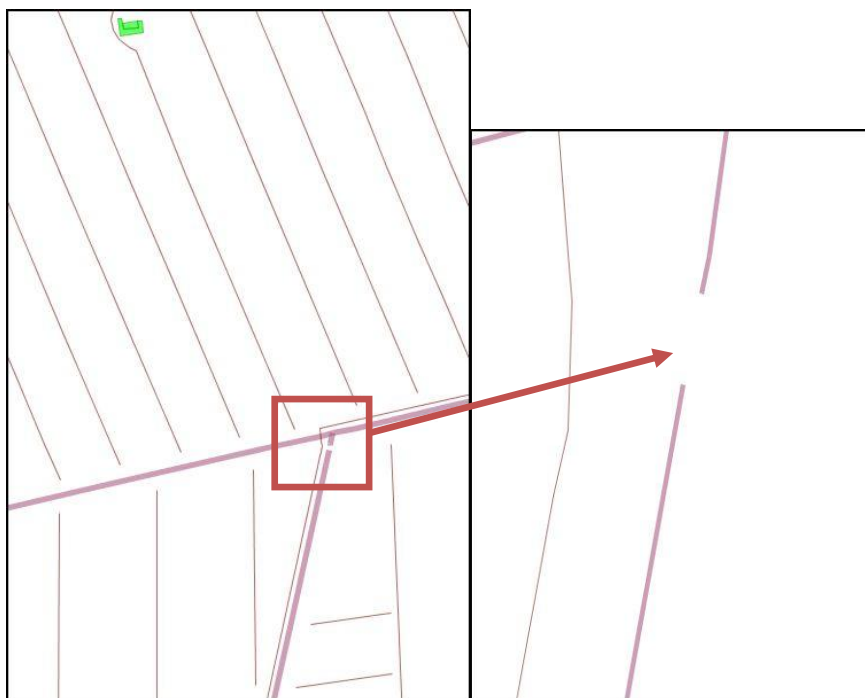


Figura 4 Esempio d'interruzione in una strada

Questa tipologia di anomalia è esplicitiva per rappresentare la differenza tra un errore e un'anomalia. Infatti due strade considerate troppo vicine potrebbero essere effettivamente così nella realtà oppure potrebbe esserci un mismatch tra i nodi o una delle due linee potrebbe arrivare "corta". Di seguito possiamo vedere un esempio di doppia anomalia cioè una situazione che genera due anomalie: una classificabile come interruzione del grafo composto dagli assi dei canali lagunari e una riguardante la forma naturale del canale troppo regolare.



Figura 5 Confronto tra un'immagine da satellite e la mappa

4.1.2 Tratto isolato o troppo corto

Questo tipo di anomalie riguarda strutture lineari isolate che rispettano la risoluzione, ma sono inferiori a un valore fissato secondo logica. Un esempio chiarificatore è un tratto di strada non connesso al resto del grafo stradale.



Figura 6 Strada isolata e scollegata dal grafo stradale

4.1.3 Mancanza vertice comune incrocio

Quest'anomalia tratta i casi in cui nel grafo dei tratti si sovrappongono senza la presenza di un punto in comune. Affligge principalmente il grafo stradale un classico esempio è un'incrocio a T. In questo caso le possibili cause dell'anomalia sono:

- Le due strade vengono a sovrapporsi perché una delle due è rappresentata più lunga del vero, e quindi non dovrebbe esserci l'incrocio.
- Potrebbe mancare effettivamente il punto che identifica l'incrocio.
- Una delle due strade potrebbe essere sopra o sotto l'altra e mancare questa informazione.

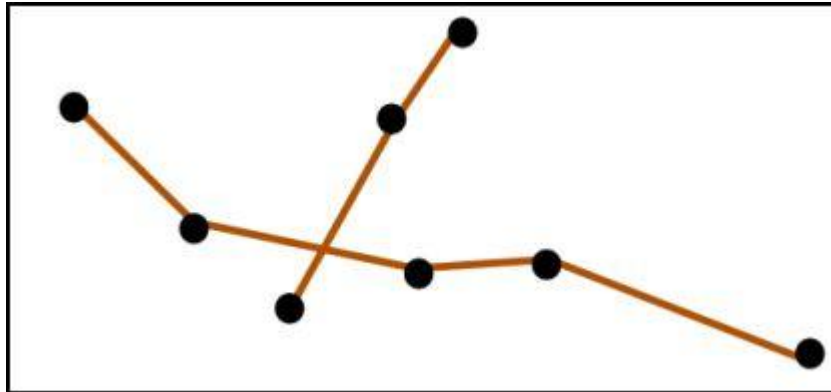


Figura 7 Zoom in un incrocio privo di punto in comune tra le due strade

4.1.4 Incoerenza di tratti adiacenti

Questa categoria rappresenta i bruschi e brevi cambiamenti del tipo di dato e della relativa vestizione nella mappa. Un esempio è una strada che da locale diventa autostrada per un breve tratto e poi nuovamente locale questo può essere dovuto a errori di codifica di un tratto.

4.2 Forma

Questa categoria di anomalie si occupa delle forme e le dimensioni degli elementi. Di seguito saranno esposti alcuni casi rilevati.

4.2.1 Limiti di dimensione legati al valore semantico

Questo tipo di anomalie si applica alle geometrie lineari e areali. Lo scopo è valutare la dimensione in base a cos'è l'elemento, cioè il suo valore semantico. La seguente figura è stata scelta per mostrare come delle anomalie possono essere la modellazione corretta di realtà particolari. Nell'immagine sono d'interesse soprattutto i due gruppi d'alberi classificati come boschi. Quello piccolo in alto a destra farà sorgere l'anomalia sopra citata a causa delle dimensioni esigue mentre il bosco al centro a causa della forma porterà alla rilevazione di un'anomalia relativa a forme naturali troppo regolari.



Figura 8 Anomalia nella forma del bosco. Confronto che mostra che rappresenta fedelmente la realtà.

4.2.2 Forme artificiali troppo irregolari

Un esempio di un caso che fa parte di questa categoria è una pista da go-kart che è classificata come impianto sportivo. Questa è rilevata come anomalia per la particolare forma tortuosa atipica per un impianto sportivo

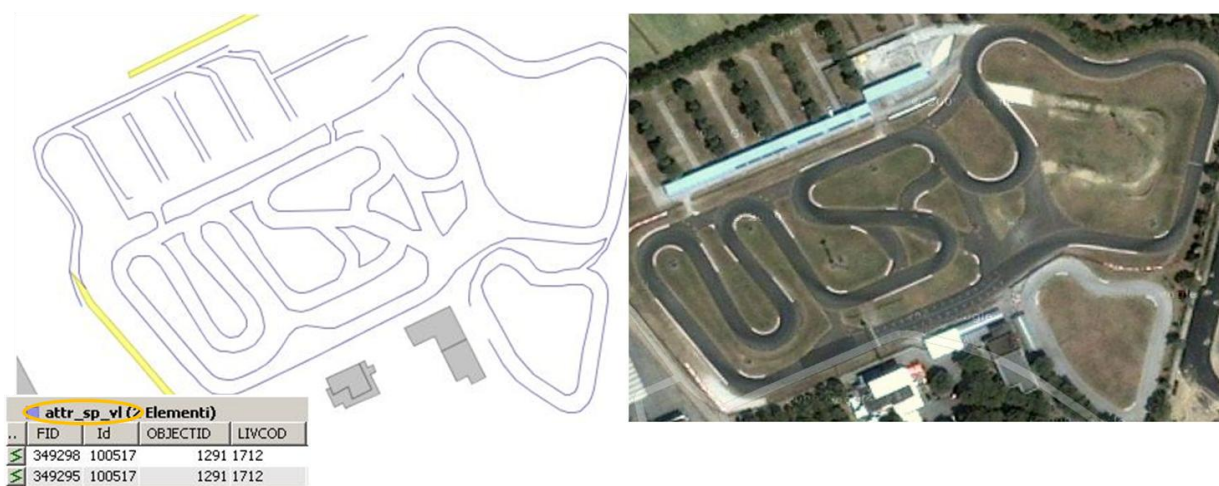


Figura 9 Pista da go-kart rilevata come figura anomala.

4.2.3 Forme naturali troppo regolari

Contiene anomalie riguardanti forme naturali e quindi laghi, fiumi, boschi e altro. Nello specifico quando un elemento naturale possiede una forma troppo regolare e squadrata. Nell'esempio sottostante Figura dieci sono rappresentati

due laghi di forma molto particolare questo è dovuto al fatto che questi sono dei laghetti artificiali per la pesca. Questa informazione non è però ricavabile dalla carta o dai dati nel database in più non sussiste alcun tipo di errore perché di per sé questo elemento è formalmente corretto ed è quindi giusto considerare questo fatto un'anomalia da segnalare.

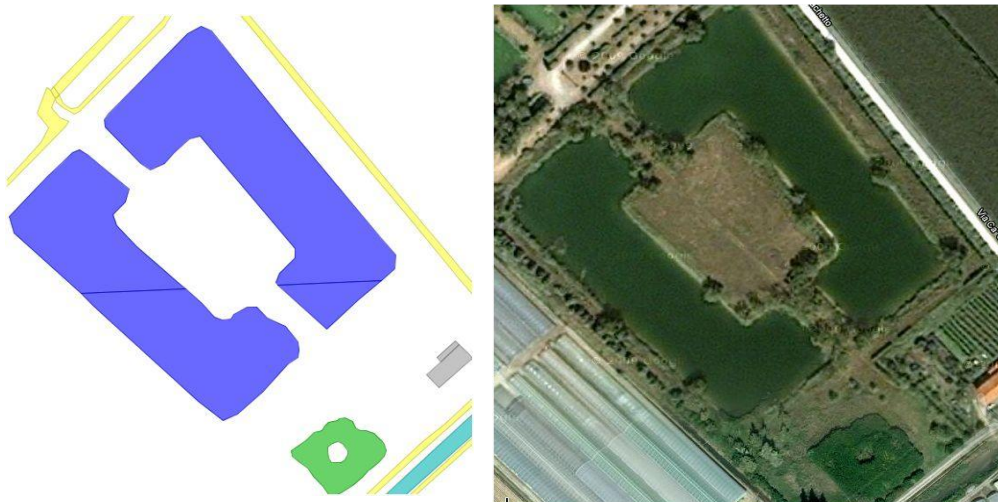


Figura 10 Anomali nella forma di questi due laghi

Nella Figura undici è rappresentato un altro esempio, questo riguarda un'isola a forma di ottagono presente nella laguna veneziana. In questo caso l'anomalia rileva un vero errore, errore di classificazione del dato nello specifico in quanto esso è registrato come specchio d'acqua cosa che come si può vedere non è.



Figura 11 Struttura anomala. Quest'ottagono è indicato come specchio d'acqua, cosa ovviamente non vera.

Un caso particolare è poi quello dei canali lagunari. Molto spesso sussistono errori topologici ma anche una volta risolti questi restano forme troncate e i grafi degli assi non connessi. Quindi queste strutture presentano molte anomalie dovute alla loro particolarità e alla difficoltà nel mapparle correttamente.



Figura 12 Laguna della zona di Chioggia. Dall'alto foto satellitare, DBT.

4.2.4 Cappi

Questa categoria rappresenta i cappi che si formano nelle strutture lineari o sui bordi di strutture areali. Questi cappi possono essere anche molto piccoli e non visibili nel risultato cartaceo ma in alcuni casi portano al fallimento degli strumenti di analisi topologici.

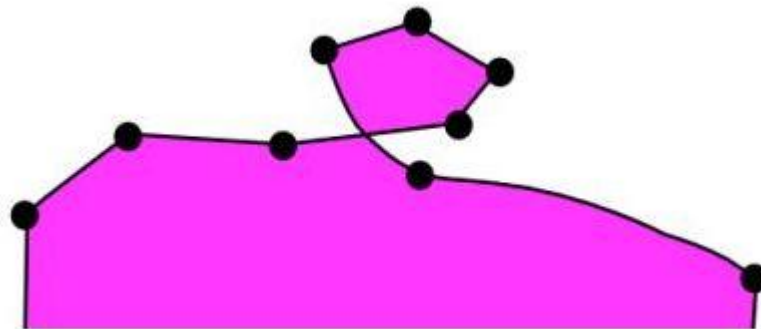


Figura 13 Esempio di possibile cappio nel bordo di un'area

4.2.5 Fiumi non monotoni

Il corso dei fiumi procede dal basso verso l'alto e dunque le quote dei vertici devono diminuire progressivamente in funzione della direzione dell'acqua. Nel caso questo non avvenga per vari motivi questo fatto provoca un'anomalia.

4.2.5 In generale relazione tra forme e posizione altimetrica

Si occupa di controlli degli angoli presenti tra le quote dei vertici di un elemento areale e verifica che questi angoli siano inferiori a un certo valore. Questo per rilevare strutture particolari come case non in piano, casi possibili nella realtà ma che potrebbe anche essere dovuti a errori negli attributi di quota dei singoli vertici.

4.3 Distribuzione

4.3.1 Distribuzioni non uniformi

Analizza la ripartizione degli elementi nella mappa, contiene sia ragionamenti sulla distribuzione totale degli elementi, cioè un semplice conteggio del numero, sia analisi sulla distribuzione non uniforme di determinati elementi. Un esempio di anomalia rilevata è se in una mappa metà contiene molti edifici e l'altra no. Ovviamente questo è possibile nella realtà in cui gli insediamenti umani hanno tipicamente forme particolari, ma è altresì vero che solitamente tendono a concentrarsi vicini tra loro dunque una zona di vuoto può essere un errore nella rappresentazione della realtà o, per esempio un parco, ma sicuramente è un'anomalia da segnalare.

4.4 Posizione

4.4.1 Posizionamento logico

Verifica che i legami logici tra il punto in cui è situato l'oggetto e il valore semantico dell'oggetto stesso siano veri. Un esempio di caso rilevato può essere la presenza di una struttura portuale in montagna o il fatto che un ponte non sia collegato a nessuna via di trasporto (strada, ferrovia ecc...).

4.4.2 Divergenza punto

Analizza se in un oggetto lineare o bordo di un'area si crea un'improvvisa divergenza tra un vertice e i due vicini a esso (il precedente e il successivo).

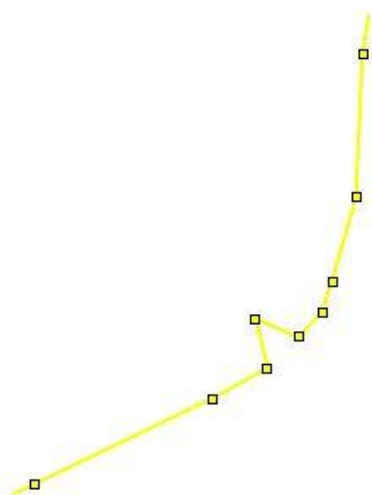


Figura 14 Esempio di divergenza di un punto appartenente a una strada

Capitolo 5 Conclusioni

In questo capitolo, che termina la tesi, saranno forniti degli algoritmi per rilevare alcune anomalie e dei possibili sviluppi futuri.

5.1 Algoritmi per la rilevazione

5.1.1 Interruzioni nei grafi

Di seguito si fornisce la descrizione di un algoritmo atto a verificare la connettività di un grafo e quindi la continuità degli elementi che lo compongono. Nello specifico quest'algoritmo si prefigge lo scopo di rilevare le interruzioni nelle strutture lineari. Per fare ciò si controlla per ogni elemento lineare se esiste un altro elemento dello stesso grafo all'interno di una certa distanza. Per fare questo si utilizza un buffer sul vertice iniziale e finale dell'elemento lineare, e non su tutta la geometria per evitare rilevazioni non volute (strade parallele vicine) e per motivi di efficienza temporale.

Dunque si allungano gli elementi lineari che hanno estremi finali o iniziali in prossimità di altri elementi lineari fino a congiungerli con questi ultimi. Si utilizzano due metodi `costruisciTabPuntiIniziali` che crea un elemento della classe `Table` contenente i vertici iniziali e `costruisciTabPuntiFinali` che crea un altro elemento della classe `Table` con i vertici finali. Si è poi realizzato un altro metodo che crea i buffer e genera gli allungamenti. Nonostante il numero di buffer raddoppi, rispetto a quelli su tutta la geometria lineare, l'efficienza aumenta essendo questi fatti su strutture puntuali, e non lineari. Il punto fondamentale dell'algoritmo è il valore del buffer: è molto importante non fissare una dimensione troppo grande in modo da non rilevare interruzioni che non sono tali ma se si utilizza un valore troppo piccolo sfuggiranno dei casi. Quest'algoritmo è da considerarsi un punto di partenza che può essere migliorato aumentando l'efficienza.

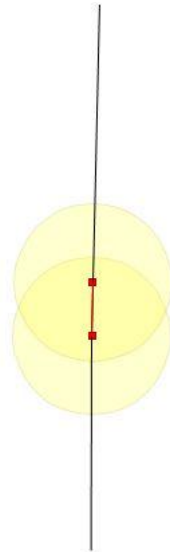


Figura 15 Output dell'algorithmo di rilevazione delle interruzioni

5.1.2 Dimensione rispetto al contenuto semantico

Per rilevare le anomalie riguardanti i limiti delle dimensioni legati al valore semantico si può procedere semplicemente fissando per ogni classe di oggetti un valore di area minimo e scandire gli elementi confrontando ogni area con il relativo valore minimo. Ad esempio se è fissato come valore minimo di area per i boschi 8 metri quadri un bosco con area di 7 metri quadri verrà segnalato come anomalia.

5.2 Conclusioni e possibili sviluppi

È stato riscontrato che la qualità dei database topografici non dipende solo dagli errori formali ma anche dalle anomalie. Infatti una delle principali tematiche della tesi è porre la questione delle anomalie e della loro rilevazione.

Gli obiettivi sono stati raggiunti per quanto riguarda fornire una tassonomia delle anomalie e in parte per quanto riguarda le indicazioni per l'implementazione.

Vista la vastità dell'argomento e delle problematiche, rimangono possibilità di ottimizzazione della tassonomia proposta e degli algoritmi di classificazione. Oltre a ciò il lavoro svolto si è focalizzato principalmente sul lato teorico e sulla presentazione del concetto di anomalia. Proprio per la vastità dell'argomento e l'approccio mantenuto si possono trarre molti spunti per degli sviluppi futuri. I principali riguardano:

- La formalizzazione della tassonomia proposta.
- Identificare parametri e metriche per la qualità globale.
- Implementare le linee guida fornite.