

UNIVERSITÀ DEGLI STUDI DI PADOVA

Facoltà di Ingegneria

Corso di Laurea Magistrale in Ingegneria Informatica

**Il processo di generalizzazione automatica
del DB50 IGM**

RELATORE
Prof. Massimo Rumor

LAUREANDO
Giovanni Langiu

CORRELATORE
Ing. Sandro Savino

Anno Accademico 2011/2012

Sommario

L'obiettivo di questo lavoro di tesi è lo sviluppo di un processo automatizzato per eseguire la generalizzazione dei dati per la scala 1:50000. Nel primo capitolo viene introdotto il tema della cartografia, definendo, in particolare, in cosa consiste il processo cartografico per la produzione delle mappe. Nel secondo capitolo viene definita la generalizzazione cartografica e descritti alcuni modelli concettuali relativi a questo tema proposti in letteratura; viene, inoltre, introdotto il tema della generalizzazione automatica e di come questo venga affrontato all'interno del progetto di ricerca Cargen. Nel terzo capitolo vengono descritte le fasi del lavoro per la messa appunto del processo e vengono descritte le caratteristiche di software sviluppato. Nel quarto capitolo viene approfondito il tema della generalizzazione semantica e come questa è stata implementata durante la messa a punto del progetto. Nel quinto capitolo vengono analizzati i risultati del lavoro svolto.

Indice

Sommario	i
1 Introduzione	1
1.1 La cartografia	1
1.2 Il concetto di scala	2
1.3 Il processo cartografico	3
1.3.1 Analisi e definizione	4
1.3.2 Acquisizione dei dati	4
1.3.3 Costruzione della mappa	5
1.3.4 Collaudo	6
1.4 GIS	6
1.5 La cartografia in Italia	8
2 La generalizzazione e il progetto Cargen	11
2.1 La generalizzazione	11
2.1.1 Generalizzazione automatica	13
2.2 Modelli di generalizzazione	14
2.3 Gli operatori della generalizzazione	16
2.3.1 Operatori geometrici	17
2.3.2 Operatori per la trasformazione degli attributi	19
2.3.3 Operatori e algoritmi	19
2.4 Il progetto Cargen	20
2.4.1 L'architettura	21
2.4.2 L'ambiente di lavoro	21

3	Il processo del DB50	23
3.1	Il processo di generalizzazione	23
3.2	Analisi preliminari	25
3.3	Le scelte iniziali	26
3.3.1	Il software	26
3.3.2	I dati di ingresso	27
3.3.3	La struttura del processo	28
3.3.4	I passi della generalizzazione geometrica	29
3.4	Caratteristiche e funzionalità del processo	31
4	Generalizzazione del modello	35
4.1	I modelli dei dati	36
4.1.1	Il DBT	36
4.1.2	Il DB25	38
4.2	Il DB50	39
4.3	La mappatura	40
4.3.1	Oggetti derivabili direttamente	41
4.3.2	Oggetti non derivabili direttamente	41
4.4	Formalizzazione delle regole e popolamento del DB50	42
5	Analisi dei risultati	45
6	Conclusioni e sviluppi futuri	49
	Appendici	53
A	Panoramica degli algoritmi di generalizzazione geometrica	53

Elenco delle figure

2.1	Esempio di generalizzazione di una carta dalla scala 1:100000 alla scala 1:200000 e 1:500000. In alto, le dimensioni delle carte sono riportate coerentemente con il valore della scala. In basso, le tre carte sono riportate secondo la stessa distanza grafica.	12
2.2	Classificazione AGENT degli operatori geometrici	18
3.1	Il modello di Gruenreich	24
3.2	Schema del processo	31
3.3	Esecuzione del processo su OpenJump. A sinistra, le tabelle del database virtuale	33
4.1	Esempio di classificazione nel modello del DBT.	38
4.2	Esempio di classificazione nel modello DB50: le strade	40
4.3	Esempio di mappatura tra i modelli DB25 e DB50: le ferrovie	41
4.4	Esempio di buffer intorno ad una tettoia per lo studio delle caratteristiche del territorio circostate	42
5.1	Generalizzazione dell'edificato: in questa serie di tre immagini si possono osservare le trasformazioni subite dall'edificato con il cambio di scala. Nell'immagine centrale si può vedere che i dati originali sono stati semplificati accorpendo edifici adiacenti o limitrofi e semplificandone la forma; solo gli edifici più piccoli sono stati eliminati. Nell'immagine a destra si può notare invece come l'insieme degli edifici sia stato radicalmente trasformato: alcuni edifici sono stati eliminati mentre altri sono stati spostati; gli edifici in genere non sono solo più regolari ma sono anche più grandi ed hanno uno spazio maggiore intorno a loro stessi.	46

5.2	Generalizzazione delle grandi aree: in ogni immagine, a diverso colore corrisponde diversa classe. Si può apprezzare come il dato si trasformi scendendo di scala: la frammentarietà del dato si riduce: i poligoni vengono accorpati tra loro od eliminati se troppo piccoli; i buchi interni ai poligoni (ad esempio i broli o le radure dentro un bosco) vengono selezionati in base alla dimensione; la forma stessa dei poligoni viene semplificata, da un lato riducendo il numero di vertici che la compongono, dall'altro eliminando repentini cambi di direzione nel profilo dei poligoni	47
5.3	Generalizzazione delle curve di livello: le immagini mostrano i punti quotati e le curve di livello. Nelle immagini è possibile osservare la riduzione del numero di punti quotati, sfoltiti sulla base della loro densità ed elevazione, e delle curve di livello, filtrate in base alla quota. Nell'immagine al centro e a destra contengono lo stesso numero di curve poiché i modelli dati rispettivamente del DB25 e del DB50 prevedono entrambi l'acquisizione di curve di livello con quota multipla di 25 metri.	47
5.4	Esempio di risultati scorretti della fase di generalizzazione geometrica dovuti a errori nei dati di ingresso	48
A.1	Nella prima immagine, la rete idrografica alla scala 1:25000: in rosso i rami individuati per il <i>pruning</i> . Nella seconda immagine il risultato della generalizzazione.	65
A.2	esempio di strada classificata erroneamente come strada principale (strada rossa).	66
A.3	Fusione di edifici adiacenti. A sinistra i dati di partenza, a destra il risultato della fusione.	66
A.4	Esempio di flusso dell'algoritmo di aggregazione di edifici.	67
A.5	Il comportamento dell'algoritmo di Sester.	67
A.6	Risultato della rimozione dei piccoli dettagli nell'algoritmo per la generalizzazione degli edifici	67
A.7	Risultato della tipificazione di un gruppo di edifici disposti a griglia.	68

Capitolo 1

Introduzione

In questo capitolo viene introdotta la cartografia e si spiega l'importanza nel descrivere aspetti del mondo reale.

Viene presentato il concetto di scala, evidenziando il ruolo che questo assume nel processo produttivo di una mappa, processo che negli anni ha avuto un'importante evoluzione con l'avvento dell'informatica.

Infine, vengono identificati i vantaggi di un'automatizzazione del processo di derivazione di una mappa a piccola scala da una a grande scala. Questa operazione, chiamata generalizzazione, è il tema principale di questa tesi.

1.1 La cartografia

La cartografia (dal greco *Chartis* = mappa e *Graphein* = scrittura) è la scienza che si occupa della progettazione, produzione, diffusione, valorizzazione e studio delle mappe. Combinando scienza, estetica e tecnica, la cartografia si basa sul presupposto che la realtà può essere modellata in modo tale da comunicare informazioni spaziali. La mappa è una rappresentazione grafica, riprodotta in scala e secondo simboli convenzionali, di una parte o di tutta la superficie terrestre, o di alcuni suoi aspetti particolari.

L'uomo ha sempre avuto la necessità di rappresentare la realtà che lo circonda e di studiare le relazioni spaziali tra specifici elementi. A tal proposito sono stati sviluppati diversi tipi di mappe, ognuno per un diverso scopo. In base alla quantità di informazioni e a finalità della mappa, se ne possono distinguere principalmente due categorie: la prima è quella delle *carte topografiche*, che forniscono informazioni metriche e descrittive di interesse generale come, per

esempio, le principali vie di trasporto e le aree urbane, la seconda è quella delle carte tematiche, che forniscono informazioni specifiche su fenomeni circoscritti e di particolare interesse.

Questa tesi tratta le carte topografiche che possono essere utili per avere una visione generale di un determinato territorio, dalla singola città alla mappa del mondo, a seconda della scala di riduzione utilizzata.

Di seguito viene discusso il concetto di scala.

1.2 Il concetto di scala

Uno degli aspetti centrali nella definizione di mappa cartografica, se non il più importante, è quello di scala, definita come il rapporto tra le misure reali e quelle della rappresentazione cartografica.

La scala di una mappa può essere espressa come una frazione o come un rapporto tra un'unità di lunghezza misurata sulla carta e il valore della lunghezza della stessa misura nella realtà, espressa nella stessa unità di misura: una misura di 1 mm di lunghezza in una mappa in scala 1:25000 corrisponde ad una lunghezza di 25 metri nella realtà.

Pensando alla scala come al valore di una frazione, che diminuisce all'aumentare del denominatore della frazione, si può dire che una mappa in 1:5000 ha una scala più grande di una mappa in 1:25000; il valore della scala è quindi inversamente proporzionale al fattore di riduzione adottato: all'aumentare del secondo diminuisce il primo.

La scala governa un gran numero dei parametri che entrano in gioco durante la creazione di una mappa, in primis il livello di dettaglio: non bisogna dimenticare che lo scopo di una mappa è la comunicazione di informazioni spaziali attraverso una rappresentazione grafica, e a seconda della scala cambia la grandezza del più piccolo dettaglio percepibile, vincolo fondamentale per ogni rappresentazione grafica.

In una mappa in scala 1:25000 una linea tracciata con un segno largo 0,5 mm copre più di 12 metri e assumendo 0,1 mm come la distanza minima che l'occhio umano è in grado di percepire, è facile calcolare che la posizione di un oggetto su una carta a tale scala, per quanto sia stata ricostruita accuratamente, sarà soggetta ad un errore di 2,5 metri. Ogni carta è caratterizzata, da un punto di vista metrico, da due parametri che sono strettamente collegati alla scala: il grado di risoluzione e l'errore massimo di posizionamento. Il grado di risoluzione, e cioè la dimensione lineare del particolare più piccolo rappresentabile, è dato dal minimo spessore del tratto grafico con cui la carta viene disegnata, e viene assunto, per convenzione, uguale a 0,2 mm; l'errore massimo di posizionamento di un punto rappresenta il diametro del cerchio al cui interno il

punto è sicuramente contenuto e corrisponde all'incertezza con cui è rappresentata la posizione di un generico punto sulla carta. Questo valore è assunto a 0,5 mm.

Il livello di dettaglio di una mappa ha un impatto sull'accuratezza dei dati presentati, sul suo contenuto e in generale sulla rappresentazione grafica, ma anche sul costo di produzione e di manutenzione della mappa stessa. Il concetto di scala ha un ruolo così importante che la produzione di una nuova cartografia è il più delle volte giustificata proprio dalla necessità di presentare dati geografici già disponibili ad un rapporto di riduzione differente. Il livello di dettaglio di una mappa influenza la precisione con cui i dati sono riportati su di essa ed anche la quantità di informazione che è possibile inserire nella mappa: la densità di informazione è infatti funzione della scala adottata. Sia il modello semantico sia la rappresentazione grafica degli oggetti devono essere calibrati sul rapporto di scala, adattandoli alla quantità di informazione che il livello di dettaglio massimo rende possibile veicolare; inoltre, sempre sulla base della scala, cambia il modo di rappresentare gli oggetti spaziali e la tipologia dei segni convenzionali usati.

A seconda della scala scelta cambia anche l'entità delle approssimazioni introdotte sui dati, sia a causa del rapporto di riduzione tra le dimensioni reali degli oggetti e la loro corrispondente rappresentazione, sia per le proiezioni adottate per ovviare al problema della curvatura della superficie terrestre.

1.3 Il processo cartografico

Con processo cartografico si definisce tutto l'insieme di procedimenti e operazioni necessari alla creazione di una carta geografica. Nonostante nel corso dei secoli il progresso delle scienze e delle tecnologie abbia portato ad uno sviluppo delle tecniche e degli strumenti in possesso dei cartografi, il processo di produzione cartografica è rimasto sostanzialmente invariato nel tempo, e può essere schematizzato nei passi seguenti:

1. Analisi e definizione
2. Raccolta dei dati
3. Costruzione della mappa
4. Collaudo

1.3.1 Analisi e definizione

Il primo passo del processo riguarda la progettazione della mappa: durante la fase di definizione ed analisi, si decidono le caratteristiche che il prodotto finito dovrà possedere. I parametri tra cui scegliere sono molteplici, quali ad esempio la scala del prodotto finito, la proiezione da utilizzare e non da ultimo la sua destinazione: se si tratta di una mappa tradizionale a supporto cartaceo, oppure una carta numerica, destinata ad una fruizione digitale.

Una delle prime mosse del cartografo è quindi quella di identificare lo scopo per la quale la mappa deve essere creata e da chi verrà utilizzata, decidendo cosa rappresentare e come rappresentarlo: le scelte effettuate in questa fase influenzeranno sia le caratteristiche tecniche della carta, sia quelle semantiche, relative cioè ai suoi contenuti.

Per quanto riguarda le caratteristiche tecniche di una carta geografica, possiamo citare la superficie di riferimento e di proiezione adottata, il tipo di rappresentazione (conforme, equivalente, equidistante) e la tipologia della sua generazione (geometrica, semigeometrica, analitica) e non da ultimo la scala.

Dal lato delle caratteristiche semantiche, in base al prodotto finale che si vuole ottenere è possibile condurre l'attività di astrazione e semplificazione dei dati in ingresso in modo tale da focalizzare l'attenzione su solo alcuni particolari aspetti della realtà da rappresentare, oppure cercare di avere una rappresentazione il più completa possibile: nel primo caso parleremo di cartografia meristica, o tematica, contenente un elevato livello di informazione su uno o più specifici argomenti, detti tematismi, nel secondo di cartografia olistica, o analitica.

E' importante sottolineare come le decisioni prese in questa fase corrispondono a definire una prima astrazione della realtà: soprattutto a livello semantico, la scelta di come distinguere e classificare gli elementi che si trovano sul territorio coincide col creare un modello di realtà. Nella produzione delle tradizionali mappe cartacee, questo processo di modellazione porta alla definizione della legenda della carta, che stabilisce quali elementi saranno presenti e come saranno rappresentati, mentre nella produzione di mappe digitali, questa fase porta alla definizione di un GeoDB.

1.3.2 Acquisizione dei dati

La fase successiva alla definizione e analisi, è la raccolta dei dati.

Esistono diversi metodi di acquisizione di dati geospaziali per i moderni sistemi informativi geografici, classificabili nei seguenti tipi [Kraak e Ormeling,2003]:

- *Rilevazioni a terra.* I dati topografici per le grandi scale possono essere acquisiti attraverso rilevazioni a terra. Spesso, i dati raccolti possono essere memorizzati in formato digitale e importati direttamente nelle comuni applicazioni GIS.
- *Rilevazioni fotogrammetriche.* La fotogrammetria è una tecnica di rilevazione della posizione di punti mediante l'utilizzo di immagini fotografiche stereoscopiche del terreno. Si tratta per lo più di immagini fotografiche riprese da aereo, in sequenze chiamate strisciate, o strip, utilizzate a coppie, e tali che ciascun fotogramma si sovrapponga per circa il 60% con quelli adiacenti e ciascuna strisciata si sovrapponga per il 15% con quelle adiacenti. Una volta ottenute le immagini, le coordinate degli oggetti possono essere determinate mediante stereoplotters analogici o, sempre più frequentemente, digitali.
Una nuova forma di raccolta dati con rilevazioni aeree è quella che utilizza il *laser altimetrico*. Per la costruzione dei modelli del terreno gli aerei sono dotati di ricevitori GPS, che permettono di determinare la rotta tracciata dall'aereo con un margine di errore di 10 cm, e di un laser (Laser Range Finder) capace di misurare distanze con precisione teorica di 1 cm. In questo modo, essendo noti la posizione dell'aereo e il tempo impiegato dal raggio laser per colpire il suolo e raggiungere nuovamente l'aereo dopo essere stato riflesso, la posizione e l'altezza dei punti del terreno possono essere determinati con un'accuratezza che dipende dall'altezza del volo e dalle caratteristiche del terreno. Attualmente queste rilevazioni hanno un'accuratezza di ± 20 cm.
- *Digitalizzazione o scansione di mappe analogiche.* Questo tipo di raccolta di dati prevede di tracciare manualmente con un cursore una sequenza di punti appartenenti a delle linee su una mappa, registrandone in formato digitale le coordinate.

1.3.3 Costruzione della mappa

Una volta raccolti i dati, si passa alla fase di costruzione della mappa: qui il cartografo, sfruttando gli strumenti a disposizione e la sua conoscenza ed esperienza, produce una carta che deve soddisfare non solo tutte le specifiche decise nella fase di definizione ed analisi, ma anche requisiti irrinunciabili per una mappa quali la leggibilità e l'usabilità. Per realizzare questo, il cartografo deve estrarre dai dati di partenza una loro rappresentazione astratta ma al contempo efficace e rappresentativa: il suo lavoro è cioè mirato alla creazione di un'astrazione della realtà geografica che faciliti la comprensione e la comunicazione dell'informazione. Con queste premesse, è possibile definire questo processo come un processo di "generalizzazione".

Parlando di generalizzazione, in ambito cartografico, ci si può scontrare con una certa ambiguità nell'uso di questo termine: nel creare una rappresentazione cartografica della realtà sono infatti coinvolte due attività di astrazione che riguardano aspetti diversi del dato geografico: da una parte il contenuto semantico, dall'altra l'informazione spaziale (posizione, forma). La definizione di generalizzazione precedente accorpa entrambe queste due attività, conosciute anche come generalizzazione del modello e generalizzazione cartografica, ma durante la fase di costruzione della mappa viene eseguito il solo processo di generalizzazione cartografica, quando il cartografo sceglie e posiziona uno ad uno gli oggetti sulla carta finale.

Il processo di generalizzazione del modello invece viene posto in essere già durante la fase di analisi e definizione delle specifiche della mappa. Nei capitoli successivi di questa tesi verrà trattato in modo esaustivo il tema della generalizzazione. La costruzione di una carta geografica include anche una fase di raffinamento estetico che si può considerare non completata durante il solo processo di generalizzazione cartografica: questa fase include attività quali il posizionamento della toponomastica, la creazione di ombreggiature o sfumature per la delineazione dell'orografia, il posizionamento di riferimenti come la griglia di inquadramento geografico e in generale la vestizione dei particolari.

1.3.4 Collaudo

L'ultima fase del processo, quella di collaudo, prevede che la carta venga sottoposta ad una serie di test per verificarne la correttezza e la consistenza; durante questa fase si può assistere ad un raffinamento estetico del prodotto, oltre che ad una verifica della validità della rappresentazione creata. Tra le attività svolte durante questa fase di collaudo si può citare il controllo della ripresa aerea, della determinazione della rete di raffittamento e dei punti d'appoggio, della restituzione, del disegno e della ricognizione e il controllo finale sul terreno mediante operazioni di misura e verifica della rappresentazione cartografica. Questa ultima è una delle attività più importanti della fase di collaudo e viene eseguita confrontando sul campo i dati riportati nella mappa con dati rilevati tramite strumenti ad alta precisione, come il GPS differenziale.

1.4 GIS

Come visto in precedenza, l'introduzione dell'informatica nel mondo della cartografia, ha introdotto notevoli cambiamenti nel modo di produrre una mappa, e sicuramente i GIS hanno contribuito notevolmente in questa direzione.

I GIS sono sistemi informativi dedicati allo studio e alla gestione di dati geografici, e attraverso questi strumenti è possibile raccogliere, modellare, manipolare, analizzare e presentare dati georeferenziati. L'acquisizione dei dati avviene normalmente scannerizzando le mappe cartacee tradizionali; un altro metodo, invece, si basa sull'utilizzo di contenitori che contengono le coordinate degli elementi geografici, espresse secondo un determinato sistema di riferimento. In entrambi i casi, l'acquisizione è seguita da un processo di astrazione e generalizzazione. Consentono, perciò, di sovrapporre differenti livelli di informazione relativi ad un'area, ottenendo una migliore comprensione dei fattori che la caratterizzano. Sono, pertanto, uno strumento completo adatto alla rappresentazione del territorio e al trattamento delle informazioni associate agli oggetti georeferenziati.

I dati spaziali, che rappresentano l'informazione territoriale codificata in un sistema informativo geografico, sono tradizionalmente divisi in due classi: raster e vettoriale. Nella prima gli oggetti sono definiti quasi ovunque nel dominio di interesse, e molto spesso sono continui. Sono rappresentati in forma discreta con matrici regolari di attributi (modelli matriciali o raster georeferenziato), reti irregolari di triangoli (modello TIN) oppure con curve di livello. La seconda classe, invece, presenta oggetti discreti e discontinui, delimitati in modo preciso. Sono rappresentati con livelli vettoriali, eventualmente topologici e tabelle di attributi associate: agli oggetti (punti, linee ed aree) sono assegnate le coordinate spaziali congruenti con la base cartografica di riferimento.

In particolare, le primitive vettoriali che costituiscono il modello sono:

1. il punto, definito da una coppia (x,y) o da una terna di coordinate (x, y, z) ;
2. la linea, costituita da una sequenza ordinata di punti;
3. il poligono, anch'esso costituito da una sequenza ordinata di punti, in cui l'ultimo punto della sequenza corrisponde al primo.

Rispetto alla rappresentazione puramente geometrica degli oggetti ad un GIS viene richiesto di mantenere e gestire le informazioni che riguardano le mutue relazioni spaziali tra i diversi elementi, cioè di strutturare i dati definendone anche la topologia. Oltre all'aspetto geometrico e topologico il GIS deve prevedere anche l'inserimento dei dati descrittivi dei singoli oggetti reali, ovvero degli attributi.

Una volta che i dati vengono memorizzati, è possibile usufruire delle funzionalità di analisi ed elaborazione degli elementi geografici degli attributi fornite dai GIS, quali:

- L'overlay topologico: in cui si effettua una sovrapposizione tra gli elementi dei due temi per creare un nuovo tematismo (ad esempio per sovrapporre il tema dei confini di

un parco con i confini dei comuni per determinare le superfici di competenza di ogni amministrazione o la percentuale di area comunale protetta).

- Le query spaziali, ovvero delle interrogazioni di basi di dati a partire da criteri spaziali (vicinanza, inclusione, sovrapposizione etc.).
- Il buffering: da un tema puntuale, lineare o poligonale definire un poligono di rispetto ad una distanza fissa o variabile in funzione degli attributi dell'elemento.
- La segmentazione: algoritmi di solito applicati su temi lineari per determinare un punto ad una determinata lunghezza dall'inizio del tema.
- La network analysis: algoritmi che da una rete di elementi lineari (es. rete stradale) determinano i percorsi minimi tra due punti.
- La spatial analysis: algoritmi che utilizzando modelli dati raster effettuano analisi spaziali di varia tipologia, ad es: analisi di visibilità.
- Analisi geostatistiche: algoritmi di analisi della correlazione spaziale di variabili georeferite.

Nei GIS le coordinate di un oggetto non sono memorizzate rispetto ad un sistema di riferimento arbitrario o al sistema di coordinate della periferica usata, ma sono memorizzate secondo le coordinate del sistema di riferimento in cui realmente è situato.

La scala di rappresentazione diventa solo un parametro per definire il grado di accuratezza e la risoluzione delle informazioni grafiche. A seconda delle scale elementi piccoli possono non essere visualizzati e si vedono solamente aree di terreno caratterizzate da una stessa quantità.

1.5 La cartografia in Italia

In Italia, la prima agenzia ad occuparsi di cartografia è stata l'IGM, Istituto Geografico Militare, il ramo dell'esercito che dal 1960 svolge le funzioni di Ente Cartografico dello Stato. Con una legge dello stesso anno, infatti, è stato deciso che l'IGM ha il compito di fornire supporto geotopografico alle Unità e ai Comandi dell'Esercito Italiano, operando per assicurare ai numerosi utenti, sia pubblici che privati, prodotti cartografici che offrano la totale garanzia sia in termini di contenuti sia di ufficialità per la descrizione del territorio. La stessa legge assegnava il compito di produrre mappe a grande scala (1:10000 o maggiore) al catasto.

L'attività dell'ente ha avuto inizio con la compilazione della "Nuova Carta Topografica d'Italia" in scala 1:100000, seguita dalla produzione della Serie25V, prodotti cartografici che coprono l'intero territorio nazionale in scala 1:25000, e della Serie50 in scala 1:50000.

Un cambiamento alla situazione cartografica italiana è stato introdotto con una legge del 1977, che assegnava ad ogni Regione italiana il compito di produrre la mappa a grande scala per il proprio territorio. Questo si è tradotto nella creazione di carte regionali in scala 1:5000 (1:10000 per le aree meno popolate), chiamate "Carta Tecnica Regionale" (CTR) con la definizione, però, di pochi o spesso nessuno standard tra le differenti carte.

Le cose sono cambiate ulteriormente con l'avvento dell'informatica, che ha spinto gli enti e le agenzie all'adozione del formato digitale per la cartografia. In tal senso, nel 2000, l'IGM ha iniziato la produzione di una nuova linea di mappe, la Serie25DB: le maggiori novità di questa linea erano l'uso esplicito di un database geografico per memorizzare le informazioni e la definizione formale di un modello di dati. L'utilizzo di questi strumenti permetteva di impostare regole di coerenza dei dati e la definizione di regole topologiche tra le classi di features, con il conseguente raggiungimento di risultati di qualità superiore.

La sempre crescente esigenza di uno standard per il modello dei dati che favorisse una maggiore collaborazione tra le regioni italiane ha portato alla creazione di gruppo di lavoro che vede partecipare le regioni e l'IGM con l'obiettivo di definire un modello di dati nazionale. Gli sforzi hanno portato, nel Febbraio 2010, alla produzione del "Catalogo dei Dati Territoriali - Specifiche di contenuto per i DB Geotopografici": tale documento definisce un modello nazionale per le mappe a grande scala (1:5000, 1:10000) e fissa i requisiti minimi che ogni cartografia regionale dovrebbe soddisfare. Una volta definitivamente approvato e adottato dalle Regioni, questo modello di dati costituirà la base per una facile condivisione dei dati geografici tra le diverse regioni d'Italia.

Capitolo 2

La generalizzazione e il progetto Cargen

Come si è potuto capire dal capitolo precedente, la generalizzazione è uno dei processi chiave nella realizzazione di una mappa.

Il progresso tecnologico degli ultimi 30 anni ha incoraggiato la ricerca di una procedura automatica che potesse prima affiancare il cartografo nel suo lavoro manuale e infine sostituire completamente il suo operato. Nonostante tutti gli sforzi e i traguardi raggiunti, al momento, quello della generalizzazione automatica è un argomento oggetto di studi.

In questo capitolo verrà introdotta la generalizzazione cartografica. Si discuterà dell'importanza di poter usufruire di un processo automatico di generalizzazione e come il progetto di ricerca Cargen, nel cui contesto questo lavoro di tesi si inserisce, abbia cercato una soluzione a tale problema. Nei capitoli successivi saranno analizzate in dettaglio le fasi di un processo di generalizzazione e come queste sono state implementate per la derivazione del DB50 IGM.

2.1 La generalizzazione

Osservando una carta geografica, si può notare come gli elementi della realtà vengono classificati, ridefiniti e posizionati, sempre secondo un ordine logico e grafico. Se non fossero compiute queste azioni di classificazione e ridefinizione, l'informazione veicolata tramite la carta sarebbe sensibilmente inferiore. Ad esempio, se tentassimo di riportare in una mappa topologica di una città in scala 1:100 000 l'intero edificato, otterremmo nient'altro che una nuvola di punti, creando solamente disordine e confusione. Se, invece, volessimo riportare nella stessa carta tutte le strade, non riusciremmo più ad individuare con facilità le strade principali, come quelle che portano da un lato della città all'altro. Ne consegue che non ha senso tentare di rappresentare nella mappa

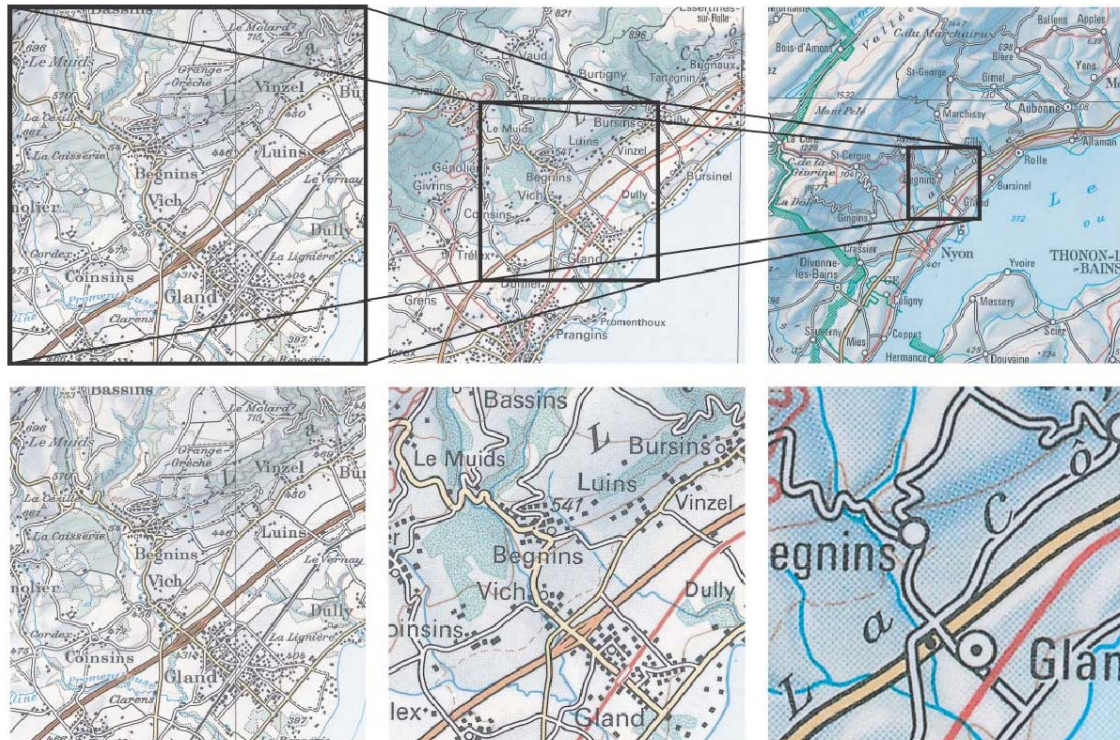


Figura 2.1: Esempio di generalizzazione di una carta dalla scala 1:100000 alla scala 1:200000 e 1:500000. In alto, le dimensioni delle carte sono riportate coerentemente con il valore della scala. In basso, le tre carte sono riportate secondo la stessa distanza grafica.

tutta la realtà. Al contrario, è necessario selezionare le informazioni rilevanti e il risultato sarà, dunque, una parte astratta della realtà.

Lo scopo della generalizzazione, così, è proprio quello di astrarre la realtà per rappresentarla in modo che essa risulti chiara, comprensibile e leggibile, tenendo presente la funzione che deve svolgere la mappa, e in accordo con la scala di rappresentazione scelta. La generalizzazione può essere quindi definita come il processo che punta alla semplificazione delle informazioni geografiche per soddisfare i vincoli di rappresentazione e per raggiungere la finalità della carta, mettendo in evidenza gli elementi importanti e al tempo stesso rimuovendo quelli irrilevanti. I fattori che devono essere tenuti in considerazione dal processo di generalizzazione sono molteplici [McMaster e Shea, 1992]:

1. *Riduzione della complessità:* il passaggio ad una scala inferiore porta diversi oggetti ad entrare in conflitto per lo stesso spazio. Ridurre il numero di questi oggetti consente di far risaltare gli elementi più importanti e di mantenere la mappa leggibile.

2. *Mantenimento dell'accuratezza spaziale*: tanto più la scala è alta e tanto più l'accuratezza spaziale deve essere rispettata, limitando al minimo l'errore dovuto alla diversa posizione degli oggetti nella mappa rispetto alla realtà.
3. *Mantenimento dell'accuratezza degli attributi* : minimizzare le alterazioni non intenzionali degli attributi delle feature (in particolare nelle mappe tematiche).
4. *Mantenimento della qualità estetica*: dai colori utilizzati alla simbologia, dal bilanciamento allo stile tipografico, vari fattori influenzano l'estetica complessiva di una mappa. L'obiettivo è una mappa bella e leggibile.
5. *Mantenimento di una logica gerarchica*: elementi di una stessa categoria, ma di importanza o dimensioni differenti, devono essere differenziati, in accordo con lo scopo della mappa. Ad esempio, una città di grandi dimensioni deve risultare molto più prominente di una città di minor dimensione, mostrando una maggiore densità del grafo stradale e dell'edificato.
6. *Coerente applicazione delle regole di generalizzazione*: allo scopo di ottenere una generalizzazione imparziale e coerente, il cartografo deve determinare esattamente quali algoritmi applicare e in quale ordine e i parametri di input necessari per ottenere il risultato voluto quando si opera ad una data scala.

Due sono le componenti che caratterizzano il processo di generalizzazione: la *generalizzazione del modello*, il cui risultato è appunto un modello astratto e rappresentativo della realtà, e la *generalizzazione cartografica*, che ridefinisce le geometrie allo scopo di fornire una coerente rappresentazione grafica del modello dei dati.

2.1.1 Generalizzazione automatica

Lo sviluppo delle tecnologie informatiche sta spingendo la comunità scientifica a cercare di implementare una procedura automatica di generalizzazione, che permetterebbe di ridurre notevolmente fatica e tempo che l'operatore umano deve dedicarvi. Per capire l'importanza che riveste l'automazione del processo di produzione delle mappe geografiche, può esser utile ricorrere ad un esempio pratico. In Francia, per la produzione con i metodi tradizionali di un foglio 91 x 121 cm per le carte topografiche in scala 1:1000000, l'IGN¹ impiega mediamente 2000 ore di lavoro, di cui 1200 per la generalizzazione e 800 per il posizionamento della toponomastica. Se supponiamo che i fogli vengano realizzati in sequenza, ci vogliono circa 20 anni per realizzare

¹Institut Géographique National, <http://www.ign.fr>

le mappe relative all'intero territorio francese. Diminuendo il tempo di produzione, invece, si possono aggiornare le mappe con sempre maggior frequenza raggiungendo, naturalmente, anche un obiettivo di risparmio economico. Così, nel 2003, l'IGN ha utilizzato per la prima volta un software per la generalizzazione automatica (LAMPS2 della Laser-Scan), e ha ridotto i tempi medi per la produzione di un foglio a 150 ore per la generalizzazione e a 160 per il posizionamento dei nomi. Ecco perchè il campo della generalizzazione automatica desta sempre più interesse presso gli enti cartografici internazionali, stimolando la produzione scientifica in questo settore. La ricerca relativa all'automazione di questi processi si è focalizzata soprattutto sulla generalizzazione cartografica, in quanto sembra l'unica delle due fasi della generalizzazione per cui è possibile immaginare una forma di automazione; la generalizzazione del modello, infatti, presuppone una capacità di astrazione ed un livello di conoscenza difficilmente trasferibile ad un calcolatore. La realizzazione di un processo di generalizzazione automatico si deve comunque scontrare con alcuni ostacoli [Mackanness e Ruas, 2007]:

- *Complessità del processo di progettazione*: la mappa è un complesso mix di pattern metrici e topologici che solitamente sono molto interdipendenti. Interpretare correttamente queste forme e individuare le caratteristiche notevoli della mappa richiede una conoscenza sia cartografica che geografica.
- *Trasformazione delle informazioni dovuta ad un cambiamento di scala*: le mappe infatti, al variare della scala, mettono in luce caratteristiche geografiche diverse. Il processo di generalizzazione deve, perciò, essere in grado di estrapolare queste differenti caratteristiche in funzione della scala, a partire dalla stessa base di dati.

2.2 Modelli di generalizzazione

La generalizzazione, dagli anni '70 ad oggi, ha subito un'evoluzione che, a partire da un approccio basato su semplici algoritmi, l'ha portata ad essere definita tramite modelli matematici sempre più complessi. Le prime attività di studio sul tema della generalizzazione, infatti, si focalizzavano sullo sviluppo di singoli algoritmi che fossero in grado di risolvere specifici problemi, come ad esempio la generalizzazione di una linea o del perimetro di un'area. Dal 1980 in poi, i ricercatori hanno cominciato a dare un'impostazione più organica ai loro studi, investigando su una possibile modellazione concettuale della generalizzazione. Alla fine degli anni '70, Robinson (1978) sviluppò uno dei primi modelli concettuali, allo scopo di capire meglio il processo di generalizzazione. Egli suddivise il processo in due sotto-processi principali: la selezione, o

preprocessing, e l'effettivo processo di generalizzazione. La prima fase coinvolge l'attività di identificazione degli elementi geografici da memorizzare nel database; la seconda interessa la manipolazione geometrica e statistica degli oggetti. Quest'ultimo sotto-processo viene suddiviso ulteriormente in tre fasi:

- semplificazione, ovvero eliminazione dei dettagli geografici non necessari;
- classificazione, cioè la categorizzazione degli oggetti;
- simbolizzazione, o associazione di una codifica grafica ad ogni elemento.

. Pochi anni dopo, Bertin (1983) parlò di generalizzazione concettuale e strutturale. Nella generalizzazione concettuale è possibile cambiare la natura di un elemento, definendo un nuovo livello di concettualizzazione. Ad esempio, nella generalizzazione concettuale si può decidere di trasformare un cluster di punti in un'area. Nella generalizzazione strutturale la natura del fenomeno rimane inalterata, mentre si agisce per semplificare la sua distribuzione; ciò avviene basandosi solo sulle informazioni già presenti in fase di elaborazione. Ad esempio, il cluster di punti, invece di cambiare natura, viene generalizzato riducendo il suo numero di punti. Grande importanza riveste anche il lavoro di Nickerson e Freeman (1986), che schematizzarono la generalizzazione usando un modello suddiviso in cinque compiti:

- modifica delle feature per mezzo di quattro operatori (selezione, semplificazione, combinazione e conversione di tipo);
- ridimensionamento dei simboli;
- displacement delle feature;
- riduzione della scala;
- posizionamento dei nomi.

In questo modello, la generalizzazione della mappa alla scala finale passa per una mappa a scala intermedia, nella quale i simboli vengono prima scalati di un fattore k , per poi essere riposizionati tramite l'operazione di displacement. Un grosso contributo deriva dal lavoro di McMaster e Shea (1992), che proposero un modello concettuale globale del processo di generalizzazione digitale. Essi suddivisero il processo in tre questioni: perché generalizzare (*why*), quando (*when*) e come (*how*). Per rispondere alla domanda *why* vengono individuati degli obiettivi filosofici (elementi teorici, elementi computazionali ed elementi specifici); in merito al *when* si specifica una valutazione cartometrica delle condizioni che motivano la generalizzazione (condizioni geometriche,

misure spaziali e olistiche, controllo delle trasformazioni); alla domanda *how* viene stilato un elenco degli operatori della generalizzazione, suddivisi in relazione all'aspetto dell'oggetto che trasformano (spaziale o degli attributi). Lamy e Ruas (1999) hanno proposto recentemente un modello di generalizzazione innovativo, basato su un sistema ad agenti, implementato poi all'interno di un progetto europeo denominato AGENT3² (AGENT, 2003). Secondo gli autori, il vantaggio principale offerto da un sistema ad agenti è quello di consentire la manipolazione degli oggetti di una mappa a diversi livelli di dettaglio. Ad ogni agente viene affidato un oggetto della mappa (micro-agent), un gruppo di oggetti (meso-agent) o l'intera mappa (macro-agent) con il compito di soddisfare i vincoli di generalizzazione imposti, utilizzando le proprie funzionalità. Gli agenti interagiscono fino a quando non viene trovata una soluzione capace di soddisfare tutti i vincoli, o se questo non dovesse essere possibile, una soluzione che soddisfi la maggior parte di essi. La letteratura scientifica recente si è occupata anche di studiare i vincoli che si presentano nel passaggio da una scala grande ad una piccola; di questi vincoli e degli operatori di generalizzazione, strumenti con i quali soddisfare i vincoli, si discuterà nel paragrafo seguente.

2.3 Gli operatori della generalizzazione

Quando si deve generalizzare una carta ci sono numerosi vincoli da soddisfare; il progetto, infatti, ha delle specifiche che devono essere rispettate nella soluzione del problema di generalizzazione. Un vincolo è una specifica di progetto che deve essere rispettata nella soluzione ad un problema di generalizzazione. Tra i più importanti vincoli si possono citare i seguenti:

- vincoli di tipo grafico, legati a parametri come la dimensione, la larghezza e la distanza;
- vincoli topologici, legati a connessione, adiacenza e inclusione tra elementi;
- vincoli spaziali, legati alla conservazione delle forme e degli allineamenti;
- vincoli semantici, legati alle relazioni logiche che intercorrono tra gli oggetti e i loro insiemi;
- vincoli di disegno, legati all'aspetto grafico della carta.

Allo scopo di soddisfare questi vincoli, una grande varietà di soluzioni sono state studiate nel corso degli anni; queste soluzioni si chiamano operatori di generalizzazione. Un operatore di generalizzazione rappresenta un tipo di trasformazione spaziale che si vuole ottenere [Weibel R

²Automated GEneralization New Tecnology, <http://agent.ign.fr/>.

e Dutton, 1999]. Gli operatori sono stati sviluppati sia emulando le pratiche manuali dei cartografi, sia tramite studi puramente matematici. Ogni operatore, comunque, prende vita mediante un algoritmo di generalizzazione che lo implementa. Diverse classificazioni sono state proposte per dare un ordine logico a questi operatori, in modo da facilitare i cartografi nella scelta dell'operatore che più si presta a soddisfare un particolare vincolo. Alcuni ricercatori suddividono gli operatori in base al tipo di geometria a cui fanno riferimento (ad esempio l'operatore di semplificazione è progettato per elementi lineari, mentre l'operatore di fusione opera su oggetti areali). Una delle più importanti è quella che distingue tra operatori che trasformano le geometrie e operatori che trasformano gli attributi [McMaster e Shea, 1992].

2.3.1 Operatori geometrici

Gli operatori geometrici sono operatori che agiscono sull'aspetto grafico e topologico di uno o più oggetti geografici. Una possibile classificazione degli operatori geometrici è stata prodotta all'interno del progetto AGENT; questi sono stati classificati in base al numero di oggetti su cui agiscono.

Nella prima classe troviamo quelli che operano su un singolo elemento, e sono:

- *simplification operator*: riduce la granularità dei contorni di linee e aree, in pratica producendo una versione semplificata dell'oggetto grazie all'eliminazione della ridondanza di punti;
- *collapse operator*: trasforma oggetti areali in punti o linee risolvendo la progressiva mancanza di spazio.
- *Enhancement operator*: valorizza un oggetto in vari modi, ovvero ingrandendo la sua forma completa (*enlargement*) o una sua parte (*exaggeration*), addolcendo il suo contorno per migliorare l'estetica (*smoothing*), squadrando la sua geometria se si avvicina ad una forma rettangolare (*squaring*);

Nella seconda classe vi sono gli operatori che possono essere applicati indistintamente ad uno o più oggetti:

- *selection operator*: seleziona gli elementi importanti considerando lo scopo della mappa;
- *elimination operator*: elimina dalla mappa gli oggetti ritenuti non importanti o ridondanti;
- *displacement operator*: sposta di posizione un oggetto o un gruppo mantenendone inalterata la forma.

Operatori geometrici della generalizzazione (Progetto AGENT)	Individuati	Simplification	Riduce la granularità dei contorni di linee e aree, mantenendo un numero minore di punti per rappresentare l'oggetto.		
		Collapse	Trasforma un oggetto N-dimensionale in uno a N-1 o N-2 dimensioni. Es: oggetti areali in punti o linee.		
		Enhancement	Enlargement	Aumento delle dimensioni di un oggetto mantenendo le proporzioni	
			Exaggeration	Ingrandimento di una parte dell'oggetto	
			Smoothing	Addolcimento del contorno di una geometria	
			Squaring	Utilizzo di una forma rettangolare per rappresentare la geometria	
		Individuati o su di un insieme	Selection / Elimination	Selection	Selezione degli elementi importanti della mappa
	Elimination			Eliminazione degli oggetti ritenuti non importanti o ridondanti	
	Displacement		Spostamento di posizione di un oggetto o di un gruppo mantenendone inalterata la forma		
	Su di un insieme (Aggregation)	Amalgamation	Fusion	Fondere in un'unica geometria areale un gruppo di poligoni	
			Merge	Fondere 2 o più linee in un'unica linea, che normalmente viene posizionata a mezzera	
		Combine	Unire in un unico oggetto, elementi che precedentemente erano separati e distinti		
		Typification	Ridurre la complessità di un gruppo di oggetti attraverso la loro eliminazione, riposizionamento, allargamento o aggregazione mantenendo la disposizione tipica di quell'insieme di oggetti.		

Figura 2.2: Classificazione AGENT degli operatori geometrici

Infine, nella terza categoria ricadono gli operatori che modificano un insieme di oggetti, e che vengono definiti operatori di aggregazione:

- *combine operator*: unisce in un unico oggetto punti che precedentemente erano separati e distinti, è identificato come un operatore 0-dimensionale, in quanto agisce su geometrie zero-dimensionali;

- *merging operator*: fonde 2 o più linee in un'unica linea, che normalmente viene posizionata a mezzzeria; è un operatore 1-dimensionale, in quanto agisce su geometrie lineari;
- *Amalgamation operator*: fonde in un'unica geometria areale un gruppo di poligoni; è un operatore 2-dimensionale, in quanto agisce su geometrie areali;
- *typification operator*: riduce la complessità di un gruppo di oggetti attraverso la loro eliminazione, riposizionamento, allargamento o aggregazione mantenendo la disposizione tipica di quell'insieme di oggetti.

2.3.2 Operatori per la trasformazione degli attributi

Gli operatori di trasformazione degli attributi si differenziano da quelli geometrici in quanto non vanno a manipolare l'informazione geometrica dell'oggetto, ma agiscono sulla sua componente descrittiva. Bisogna infatti ricordare che la scala di una mappa influisce anche sulla granularità del modello di dati. Ad esempio, un municipio e un centro commerciale posso essere riclassificati come edifici generici, a causa di una riduzione di scala che mette in secondo piano l'utilizzo dell'edificio stesso. Gli operatori di questo tipo sono:

- *classification operator*: riduce i dati grezzi in un insieme di classi, operazione spesso necessaria in quanto non è sempre praticabile associare ad ogni singolo fenomeno un simbolo diverso per successivamente mapparli;
- *symbolization operator*: codifica graficamente i dati e può essere applicato sia alla componente descrittiva sia a quella geografica.

2.3.3 Operatori e algoritmi

Gli operatori appena esaminati descrivono, di fatto, le singole operazioni che il cartografo compie per generalizzare la mappa. Ad ogni operatore, poi, corrisponde un algoritmo che implementa una trasformazione del dato geografico, anche se è facile imbattersi in più implementazioni dello stesso operatore, a seconda dell'oggetto su cui agisce. Ad esempio, l'operatore di semplificazione viene realizzato in modo diverso quando si applica ad una strada rispetto al caso in cui si applichi ad un edificio. Ogni algoritmo è caratterizzato da una serie di parametri, che vanno impostati in funzione della scala, dell'oggetto a cui fanno riferimento, e in generale, del contesto. Lo stesso algoritmo di semplificazione, così, potrebbe essere applicato con parametri differenti a seconda che l'oggetto si trovi all'interno di un centro abitato, dove troppi dettagli tendono a creare

confusione, a in una zona scarsamente abitata, dove non ci sono questo tipo di complicazioni. Il primo problema da affrontare è, quindi, la scelta dei parametri degli algoritmi e l'opportunità di variarli in funzione del contesto. Il secondo problema, non meno importante, è l'ordine con cui gli algoritmi vengono eseguiti: il cartografo, per decidere la sequenza delle azioni da compiere e la loro coerenza con lo scopo della mappa o con il contesto, utilizza la sua conoscenza cartografica e la sua capacità di vedere sotto diverse astrazioni la realtà rappresentata. Gli algoritmi che implementano gli operatori di generalizzazione sono, infatti, potenti strumenti per risolvere i vincoli spaziali, semantici e grafici; essi, però, non vanno utilizzati indiscriminatamente sull'intera area da generalizzare o sempre secondo la stessa sequenza statica, ma vanno scelti, settati e applicati ogni volta diversamente, in funzione del contesto, della scala e dello scopo della mappa.

2.4 Il progetto Cargen

Il progetto CARGEN, CARTographic GENeralization, è un progetto di ricerca, partito nel 2006, che vede coinvolti il Dipartimento di Ingegneria Informatica dell'Università di Padova, la Regione Veneto, e l'Istituto Geografico Militare. L'obiettivo principale era quello di elaborare nella sua interezza (progettazione, sviluppo e test) un processo automatizzato di generalizzazione cartografica, che partisse dai dati in scala 1:5000 del modello GeoDBR - sviluppato dalla Regione Veneto - per restituire una base di dati coerente con il modello DB25 dell'IGM. Avendo ottenuto dei risultati soddisfacenti, il lavoro di ricerca fu poi esteso alla generalizzazione di mappe con scala 1:50000 e nel 2009 partì la seconda fase del progetto Cargen, su cui questa tesi si focalizza. Sebbene i dati di partenza si riferiscono appunto al più recente modello GeoDBR, in questo progetto si è fatto uso anche della CTRN³, fornita sempre dalla Regione Veneto. In particolare, la relazione che intercorre tra i modelli ha come punto di partenza la CTRN, che viene utilizzata per popolare il GeoDBR, dal quale si derivano successivamente i dati per popolare il DB25. Seppur già da tempo, in Europa, la produzione automatizzata di carte a differenti scale attraverso la generalizzazione è prassi ormai consolidata, in Italia la situazione risulta essere diametralmente opposta. Ecco perchè questo progetto, essendo tra i primi in Italia, cerchi anche di essere uno stimolo per questo tipo di produzione. Sin dall'inizio si è deciso di intraprendere la strada del raggiungimento di risultati concreti: ad innovativi approcci teorici fine a se stessi si è preferita la realizzazione di un prodotto reale, ossia un software che effettivamente realizzasse la generalizzazione cartografica.

³Carta Tecnica Regionale Numerica

2.4.1 L'architettura

Il modello adottato all'interno del progetto CARGEN per la gestione dei dati è stato quello client-server. Il server è costituito da una macchina con installato un DBMS Oracle Spatial 10g, la cui funzione è quella di memorizzare e mantenere i dati spaziali, accessibili tramite query. Nel lato client, invece, sono stati installati i software Geomedia Professional 6 e Dynamo/Dynagen, entrambi di proprietà della Intergraph. Geomedia viene utilizzato principalmente come strumento d'accesso ai dati spaziali; Dynamo/Dynagen sono invece gli strumenti usati durante il processo di generalizzazione cartografica. Questa architettura, tuttavia, è stata in parte abbandonata recentemente: infatti, sia allo scopo di semplificare lo sviluppo di nuovi algoritmi, sia allo scopo di migliorare le prestazioni temporali, è stato cambiato metodo d'accesso ai dati ed il software per visualizzarli. I dati vengono caricati in RAM e gestiti proprio come se fossero delle tabelle, grazie ad una libreria sviluppata all'interno del Progetto. Inoltre, una libreria potentissima sviluppata in Java, la JTS⁴, fornisce una vasta serie di operatori spaziali, evitando così di ricorrere al DBMS di Oracle per effettuare le interrogazioni spaziali. Nel paragrafo successivo saranno brevemente descritti i software più importanti usati durante l'attività di tesi.

2.4.2 L'ambiente di lavoro

JTS Topology Suite La JTS Topology Suite è una libreria open source, scritta interamente in Java, che fornisce una modellazione ad oggetti per le geometrie lineari in uno spazio euclideo. In questa libreria sono definite tre geometrie fondamentali, *Point*, *LineString* e *Polygon*, che rappresentano rispettivamente la geometria puntuale, lineare e areale. La JTS mette a disposizione numerose funzioni geometriche, tra le quali possiamo citare:

- gli operatori topologici, che realizzano le funzioni di intersezione, differenza, unione;
- la funzione per la creazione del buffer intorno alla geometria;
- la funzione per la costruzione dell'involuppo convesso;
- alcune funzioni per la semplificazione delle geometrie, come l'algoritmo di Douglas-Peucker;
- la funzione per la costruzione del Minimum Bounding Box.

⁴JTS Topology Suite, <http://www.vividsolutions.com/jts/jtshome.htm>

Oltre a queste funzioni, la JTS fornisce l'implementazione di indici spaziali, come il *quadtree*⁵, che offrono un modo veloce per la risoluzione di *query* spaziali.

OpenJump OpenJump è un Desktop GIS open source che permette di visualizzare, modificare e interrogare dati spaziali. OpenJump è scritto in Java, si basa sulla JTS ed è in grado di gestire file raster, vettoriali e database (PostGis, Oracle, ArcSDE); una caratteristica degna di nota è la sua architettura modulare, che permette di estendere di molto le funzionalità di base, potendo integrare, per esempio, il proprio codice mediante la realizzazione di un plugin. In OpenJump, la creazione di un plugin, relativo al proprio codice, offre al programmatore il grosso vantaggio di poter visionare tramite l'interfaccia grafica gli effetti della propria applicazione. Il plugin diventa così uno strumento essenziale nello sviluppo di nuovi algoritmi che manipolano geometrie: avere una risposta grafica e istantanea è un grande aiuto per semplificare e velocizzare la fase di testing e debug. La modalità con cui OpenJump gestisce le *feature* si basa sull'utilizzo dei layer, o livelli, che svolgono il ruolo di contenitori di *feature*. Ogni layer è in grado di contenere le *feature* relative ad uno specifico schema dati, chiamato FeatureSchema; quest'ultimo specifica il nome e la tipologia degli attributi che costituiscono la *feature*, simile a quanto accade nelle tabelle dei database. Un layer rappresenta, quindi, una vera e propria tabella, il cui schema dati è specificato dal FeatureSchema. Il layer è interrogabile per mezzo di *query*, che possono essere sia spaziali che non spaziali. Per migliorare le *query* spaziali, è possibile associare al layer uno degli indici spaziali forniti dalla JTS.

Eclipse Eclipse è un ambiente di sviluppo integrato, multi-linguaggio e multiplatforma. Ideato nel 2001 da un consorzio non-profit di grandi società quali Borland, IBM, Red Hat, SUN, Ericsson, HP, Fujitsu, Intel, MontaVista Software, QNX, SAP e Serena Software, chiamato Eclipse Foundation, viene supportato da una comunità strutturata sullo stile dell'open source. La licenza dunque di Eclipse è la Eclipse Public License, permette cioè di creare prodotti derivati ridistribuibili gratuitamente. Importante caratteristica deriva dal fatto che, essendo scritto in Java, Eclipse è multiplatforma. È disponibile infatti per le piattaforme Linux, HP-UX, AIX, Mac OS X e Windows. Dal 2006, la Eclipse Foundation ha prefissato un'uscita annuale del suo software.

⁵Struttura dati ad albero nella quale tutti i nodi interni hanno esattamente quattro nodi figli. Viene utilizzata per partizionare lo spazio bidimensionale, suddividendolo ricorsivamente in quattro quadranti.

Capitolo 3

Il processo del DB50

In questo capitolo viene descritto il modo in cui è stato progettato il processo di generalizzazione per il DB50 IGM: vengono illustrate le analisi preliminari e come queste abbiano influito sulla progettazione. Riprendendo quanto visto nel capitolo precedente, saranno discussi i concetti di generalizzazione semantica e geometrica e il modo in cui questi due aspetti della generalizzazione cartografica sono stati inseriti all'interno del processo, giustificando le scelte di progettazione.

3.1 Il processo di generalizzazione

La generalizzazione cartografica può essere scomposta in due processi distinti [Grunreich, 1985]: generalizzazione model-oriented, il cui risultato è un database geografico, e generalizzazione graphic-oriented, che produce invece una vera e propria mappa.

Nel caso del DB50, il processo di generalizzazione potrebbe appartenere alla prima tipologia, in quanto non si intende utilizzarlo direttamente per stampare una mappa. Tuttavia le specifiche IGM per il DB50 contengono direttive e requisiti per la rappresentazione dei dati: questo ha permesso di implementare all'interno del processo entrambi gli aspetti della generalizzazione, ottenendo un prodotto a metà strada tra le due classificazioni. In riferimento a quanto appena detto, nell'ambito del progetto Cargen e di conseguenza in questa tesi, si parla di rispettivamente di generalizzazione semantica e generalizzazione geometrica. La prima gestisce la traduzione dell'informazione semantica del modello di dati di partenza in quella del nuovo modello, operando una riclassificazione degli oggetti e stabilendone, di fatto, gli attributi per la memorizzazione nel nuovo database. La seconda gestisce la trasformazione dell'informazione spaziale dei dati di origine: le geometrie possono dover essere modificate sia per adattarsi al nuovo modello di dati

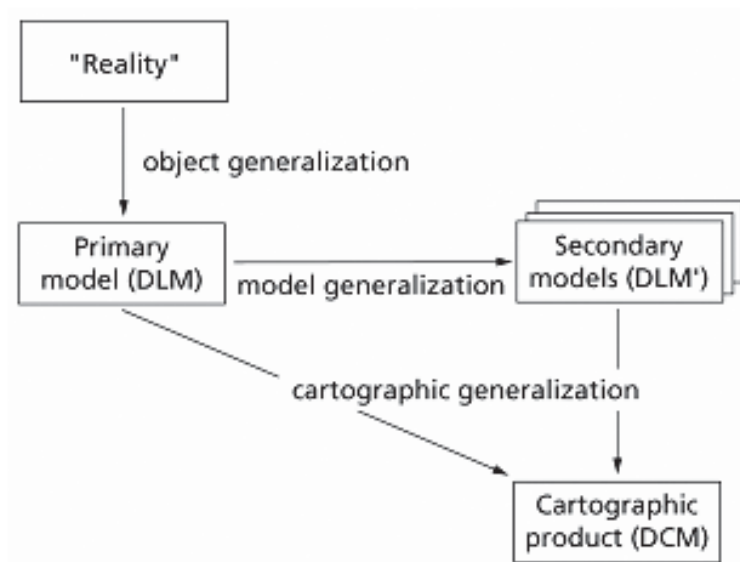


Figura 3.1: *Il modello di Gruenreich*

e le specifiche di derivazione (ad esempio, un oggetto rappresentato da un'area alla nuova scala - nel nuovo modello - ha geometria puntuale), sia per rispettare qualche specifica (per esempio, due edifici a distanza reciproca inferiore ad una certa soglia devono essere uniti). Avendo come obiettivo la messa a punto del processo di generalizzazione del DB50 IGM e potendo disporre degli algoritmi di generalizzazione geometrica precedentemente sviluppati nell'ambito del progetto Cargen, il presente lavoro di tesi si focalizza maggiormente sulla progettazione e sullo sviluppo del processo di generalizzazione semantica. Dopo aver progettato e sviluppato il nuovo modelli dei dati, studiato e realizzato le procedure per la derivazione dei dati del nuovo modello dal modello sorgente e combinato i risultati è stato possibile implementare un processo completo di generalizzazione. Nei paragrafi successivi verranno analizzate nel dettaglio le fasi della progettazione e dello sviluppo, discutendo e giustificando le scelte effettuate dopo un'accurata analisi dei requisiti e delle specifiche di progetto.

3.2 Analisi preliminari

La progettazione del processo ha richiesto un'approfondita analisi preliminare della documentazione esistente, che ha fornito un'ottima guida sulla costruzione del modello dei dati e delle procedure dell'intero processo. Il primo obiettivo del lavoro è stato la creazione di un modello dei dati per la scala 1:50.000. Quando si vuole modellare una determinata realtà sono molteplici gli aspetti di cui si deve tener conto per decidere quali elementi sono ritenuti necessari e devono quindi essere contenuti nel modello e come questi devono essere rappresentati. In particolare, nell'ambito della cartografia, la scala di rappresentazione e lo scopo per cui una mappa viene creata giocano un ruolo fondamentale in questo senso. In riferimento al nostro obiettivo, il documento che maggiormente ha rappresentato un utile riferimento in tal senso è la Carta d'Italia alla scala 1:50.000 IGM. Il documento ufficiale dell'Ente descrive i particolari topografici da inserire nel modello per la Carta d'Italia e i corrispondenti segni convenzionali da impiegare per la loro rappresentazione grafica. A questo punto, avendo a disposizione il modello dei dati per il DB25, il passo successivo è stato quello di analizzare tale modello. Un importante lavoro è stato quello del confronto delle specifiche dei due modelli: il controllo della coerenza tra le specifiche individuate per il nuovo modello e quelle relative al DB25 ha permesso, di fatto, la verifica della correttezza dei risultati dell'analisi precedente. Dopo l'analisi del contenuto e della struttura del modello del DB25, infine, sono stati definiti i requisiti che il modello del DB50 avrebbe dovuto soddisfare. In particolare, il modello deve:

- Contenere tutti gli oggetti della Carta d'Italia
- Permettere la derivabilità dal DB25
- Classificare gli oggetti secondo la struttura IGM

Una volta chiare le specifiche per lo sviluppo del modello, è stato necessario reperire informazioni utili allo sviluppo del processo completo di generalizzazione. La Carta d'Italia, oltre alle specifiche del modello, fornisce importanti indicazioni per la fase di generalizzazione geometrica. Studiando le norme e i criteri per l'esecuzione delle elaborazioni grafiche in fase di derivazione dei dati dalla cartografia 1:25.000, presenti nel documento, è stato possibile definire i requisiti del processo in termini di funzionalità e algoritmi richiesti per l'esecuzione del processo di generalizzazione geometrica. Nella seguente tabella vengono riportati gli algoritmi principali e più complessi di generalizzazione, mettendo in evidenza quali procedure sono state sviluppate appositamente per la generalizzazione alla scala 1:50.000 e quali, invece, erano stati

implementati per la generalizzazione alla scala 1:25.000 e riadattati per lavorare con i dati del nuovo modello:

Algoritmo	1:25.000	1:50.000
Generalizzazione rete idrografica	x	
Generalizzazione rete stradale		x
Generalizzazione rete ferroviaria	x	
Generalizzazione edificato		x
Generalizzazione grandi aree	x	
Generalizzazione elementi lineari	x	
Generalizzazione elementi lineari	x	

Oltre a questi algoritmi principali sono stati individuati alcuni algoritmi minori per la modifica semplice delle geometrie e per la valutazione di alcuni parametri in fase di generalizzazione semantica. Per questi, come vedremo più avanti, è stato sviluppato del codice ad hoc. Una volta definiti i requisiti del processo è stato possibile effettuare varie scelte di progetto che verranno presentate di seguito.

3.3 Le scelte iniziali

La seconda fase del progetto Cargen, avente come obiettivo lo sviluppo del processo di generalizzazione del DB50 IGM, nasce come naturale continuazione della prima. Nonostante i risultati di quest'ultima siano da considerarsi il punto di partenza per raggiungere l'obiettivo, durante lo sviluppo dell'attuale progetto sono state fatte delle scelte sia di carattere tecnologico che di processo che differenziano i risultati finali delle due fasi.

3.3.1 Il software

La prima fase del progetto Cargen ha portato allo sviluppo in-house di un software realizzato in Java, basato su una architettura client-server legata al DBMS Oracle Spatial. In linea con le specifiche del progetto, il software è stato sviluppato in modo che fosse completamente automatico e non richiedesse interventi dell'utente durante l'esecuzione; il software si avvia da riga di comando ed è possibile per l'utente impostare i parametri dei vari algoritmi di generalizzazione prima dell'inizio del processo. Durante la fase di test, però, si era osservato che il software così prodotto aveva alcuni aspetti migliorabili; tra questi in particolare:

- la dipendenza da un DBMS risultava limitante sia da un punto di vista tecnico, poiché rendeva più complesso lo sviluppo, il test e l'esecuzione degli algoritmi e ne aumentava

i tempi di esecuzione a causa delle continue transazioni da e verso il database, sia da un punto di vista della flessibilità, poiché rendeva dipendente il software da un software commerciale di cui non si aveva il completo controllo;

- le librerie sviluppate per la manipolazione dei dati geometrici erano molto sensibili alla qualità dei dati di ingresso e non erano in grado di gestire geometrie complesse; questo richiedeva continui controlli sulle geometrie durante tutto il processo che appesantivano e rendevano più complesso il codice.

Durante la fase di progettazione del nuovo processo si è deciso che il nuovo software avrebbe dovuto provare a superare queste limitazioni. Grazie a nuove librerie, in parte sviluppate all'interno del laboratorio *GIRTS* del Dipartimento di Ingegneria dell'Informazione dell'Università di Padova, in parte acquisite con licenza LGPL dal pubblico dominio, è stato possibile creare un nuovo ambiente di sviluppo per gli algoritmi di generalizzazione il quale, oltre a nuove funzionalità, consente una gestione notevolmente migliorata delle primitive geometriche e permette di scrivere software che non è legato ad un DBMS specifico.

3.3.2 I dati di ingresso

La generalizzazione dei dati alla scala 1:50000 può essere ottenuta partendo dai dati originali in scala 1:5000, oppure dai dati già generalizzati alla scala 1:25000. Entrambi gli approcci hanno vantaggi e svantaggi, collegati principalmente alla quantità di informazioni presenti nei dati: alla scala 1:5000 è possibile accedere ad un maggior numero di dati, mentre alla scala 1:25000 il minor numero di dati, inoltre già generalizzati, permette una elaborazione più veloce. Inizialmente, sulla base di una serie di considerazioni sia teoriche che tecniche, la scelta era stata quella di preferire una maggiore quantità di informazione rispetto alla velocità di esecuzione e quindi di generalizzare i dati in scala 1:50000 a partire dai dati originali alla scala 1:5000. Da un punto di vista teorico il problema della generalizzazione cartografica può essere visto come un problema di estrazione di informazione: quando l'occhio umano legge una carta percepisce sia alcune informazioni che sono esplicite, tratte dalla rappresentazione dei dati (ad esempio dalla simbologia contenuta nella legenda), sia altre informazioni che sono invece implicite, ricavate dalla organizzazione dei dati (ad esempio dalla vicinanza con altri oggetti). Per poter generalizzare correttamente l'informazione geografica in modo automatico, è necessario che il software sia in grado di estrarre ed analizzare sia le informazioni esplicite, contenute nel modello dati, sia quelle implicite: è abbastanza ovvio che maggiore è la quantità di informazioni a disposizione,

tanto più accurata potrà essere la generalizzazione. Da un punto di vista tecnico, il tempo di esecuzione del processo non è stato mai considerato un obiettivo primario all'interno del progetto di ricerca e, per questo motivo, l'ipotizzabile incremento di velocità portato dall'utilizzare dati generalizzati è stato valutato meno importante della possibilità di accedere alla maggiore informazione contenuta nei dati sorgenti. Nonostante queste considerazioni, però, la scelta è caduta sulla possibilità di utilizzare come dati di input quelli generalizzati dal processo precedente. I motivi di tale scelta sono da ricercarsi soprattutto nella natura dei modelli di dati in gioco. Infatti quello del DB25 IGM è un modello ben formalizzato e adottato a livello nazionale, a differenza di quello del GeoDBR che varia in base alle scelte progettuali di ogni regione¹.

3.3.3 La struttura del processo

Nel paragrafo precedente sono stati presentati i concetti di generalizzazione semantica e geometrica. Il processo è stato progettato proprio sulla base di questa distinzione: in due step successivi vengono eseguite prima la generalizzazione semantica e poi quella geometrica. Anche per quanto riguarda la struttura del processo sono state effettuate scelte diverse rispetto a quelle del processo sviluppato nella fase precedente del progetto Cargen: in quest'ultimo, infatti, veniva applicata prima la generalizzazione geometrica e dopo quella semantica. Questa scelta era dettata dal fatto che la base di dati derivata non può contenere maggiori informazioni della base di dati originale e una tale soluzione permette agli algoritmi destinati alla modifica delle geometrie di operare con il maggior numero possibile di dati, potendo agire quindi sulla base di informazioni più dettagliate che potrebbero andare perse nel passaggio al modello per la scala più piccola. Inoltre, per essere in grado di accedere alle informazioni presenti nei dati di ingresso, gli algoritmi di generalizzazione devono conoscere come questi dati sono strutturati; nel caso del primo progetto questo richiedeva, quindi, che gli algoritmi di generalizzazione geometrica fossero sviluppati sulla struttura del GeoDBR. Questa scelta rende di fatto molto difficile applicare gli algoritmi così sviluppati a dati strutturati secondo un diverso modello, poiché questa operazione richiederebbe di modificare in tutti gli algoritmi le funzioni di accesso ai dati. Nello sviluppo del nuovo progetto si è deciso di operare in maniera diversa per ottenere un codice che si adattasse in modo più flessibile a dati di ingresso strutturati secondo modelli diversi. Questa scelta potrebbe sembrare in contraddizione con il fatto che il progetto nasce specificamente sui dati della Regione Veneto, ma è motivata dalle seguenti osservazioni:

¹Recentemente questa limitazione del GeoDBR è stata superata con l'introduzione del DBT - Database Topografico - il cui modello è unico a livello nazionale

- le definizioni dei modelli dei dati coinvolti sono cambiate ripetutamente durante lo svolgimento del progetto
- una maggiore flessibilità nella gestione dei dati di ingresso può permettere in futuro di applicare il processo sviluppato ad altri scenari.

Per ottenere questo risultato il flusso del processo è stato invertito, facendo precedere la fase di generalizzazione semantica a quelle geometrica, e gli algoritmi di quest'ultima sono stati sviluppati sul modello dei dati generalizzati. Questo permette, in caso di generalizzazione di dati di ingresso con un diverso modello, di poter lanciare il processo senza dover modificare gli algoritmi, che rappresentano la parte più rilevante e complessa del software.

3.3.4 I passi della generalizzazione geometrica

Quella delle generalizzazione geometrica è l'ultima fase del processo di generalizzazione. Essa ha lo scopo di ottenere una rappresentazione derivata non solo chiara e leggibile ma coerente con la scala della cartografia che si vuole ottenere. Infatti, la riduzione di scala di una carta è solitamente accompagnata dalla riduzione degli oggetti visibili, così come l'ampliamento e/o l'ingrandimento di oggetti considerati fondamentali che altrimenti non potrebbero essere visibili a quella scala. La generalizzazione geometrica è sicuramente la parte più complessa del processo di derivazione e richiede lo sviluppo di algoritmi più complessi. Nonostante la disponibilità di algoritmi già sviluppati, anche il processo di generalizzazione geometrica ha richiesto uno studio preliminare al fine di poter essere integrato correttamente all'interno del processo completo. Ogni algoritmo, infatti, esegue la generalizzazione di uno strato informativo ben preciso (rete idrografica, rete stradale, edificato etc.) e nel flusso del processo agisce senza alcuna interazione con gli altri. Ci sono dei casi in cui l'output di un algoritmo costituisce l'input di uno degli altri algoritmi, con i dati eventualmente elaborati e preparati appositamente. Risulta dunque chiaro che l'ordine in cui vengono eseguiti gli algoritmi assume un'importanza fondamentale per la buona riuscita della generalizzazione geometrica. Nel decidere l'ordine di esecuzione degli algoritmi, inoltre, è stato necessario ancora una volta fare riferimento anche alle specifiche IGM: un chiaro esempio di come queste influiscano sul flusso della generalizzazione geometrica è quello dell'operazione di armonizzazione. Con la riduzione della scala può capitare che due particolari topografici si sovrappongano: ne consegue la necessità di provvedere alla variazione della posizione planimetrica o alla modifica della forma per molti di essi, così da evitare sovrapposizioni grafiche e, al contempo, mantenere sulla carta derivata l'informazione riguardante la posizione relativa che gli oggetti topografici hanno sulla carta alla scala di partenza. Si rende quindi neces-

sario valutare, di volta in volta, quali particolari mantenere in posizione corretta, quali spostare e di quanto, tenendo conto del contesto cartografico limitrofo. In questo senso, l'IGM indica delle regole fondate sull'adozione di una gerarchia tra i particolari, che tiene conto della loro natura ed importanza in funzione della scala e delle finalità della carta. I particolari della rete idrografica, per esempio, occupano il primo posto tra gli oggetti che non devono essere spostati, salvo particolari condizioni cartografiche. Seguono le ferrovie, le autostrade, le strade di categoria inferiore, i fabbricati e la vegetazione etc. Confrontando quindi le indicazioni fornite dall'Ente e le caratteristiche degli algoritmi disponibili, è stato possibile organizzare il flusso del processo di generalizzazione geometrica, che viene eseguito secondo i seguenti passi:

1. generalizzazione della rete idrografica
2. fusione degli edifici
3. generalizzazione della rete stradale
4. generalizzazione della rete ferroviaria
5. generalizzazione degli edifici
6. generalizzazione dei fossi
7. generalizzazione degli elementi lineari
8. generalizzazione delle grandi aree
9. generalizzazione degli elementi puntuali

Lo strato idrografico viene considerato il più importante e viene generalizzato nel primo passo del processo, in modo da non subire l'influenza della generalizzazione di altri strati informativi. Gli edifici vengono generalizzati in due fasi. Nella prima si fondono quelli adiacenti e vengono eliminati, riducendo il numero di oggetti da elaborare nei passi successivi. La seconda, in cui vengono trattati gruppi di edifici, viene eseguita dopo la generalizzazione di tutte le reti (fiumi, strade e ferrovie) in quanto queste vengono utilizzate per partizionare lo spazio, in modo da poter elaborare in modo indipendente oggetti eventualmente vicini ma separati da un elemento della rete. Strade e ferrovie vengono generalizzate appunto nel terzo e quarto passo del processo. I fossi non fanno parte della rete idrografica e vengono quindi inseriti in un altro modulo, generalizzati ricorrendo alla tipificazione (a differenza degli elementi dell'idrografia, generalizzati tipicamente mediante selezione). Vengono generalizzati dopo gli edifici perchè l'operatore di tipificazione

utilizza la posizione di questi per creare le geometrie tipificate. Gli elementi lineari, tipicamente recinzioni e muri, devono essere generalizzati dopo gli edifici perchè vengono selezionati e modificati in base al contenuto dell'area che circondano. Le grandi aree (boschi, laghi, terreni coltivati..)devono essere estese fino ai limiti di strade, fiumi o recinzioni e vengono dunque generalizzate dopo queste classi. L'ultimo passo del processo è la generalizzazione degli elementi puntuali quali come alberi e quote altimetriche, che subiscono principalmente uno sfoltimento in base alla loro posizione relativa agli altri elementi dello spazio. Per una breve descrizione degli algoritmi citati si rimanda all'Appendice.

3.4 Caratteristiche e funzionalità del processo

Il prodotto finale del lavoro, ovvero il software che implementa il processo, è organizzato in moduli eseguiti in sequenza, ognuno dei quali rappresenta una fase del processo. L'intero processo è stato implementato utilizzando il linguaggio Java. Il modulo di *preprocessing*, come spiegato in

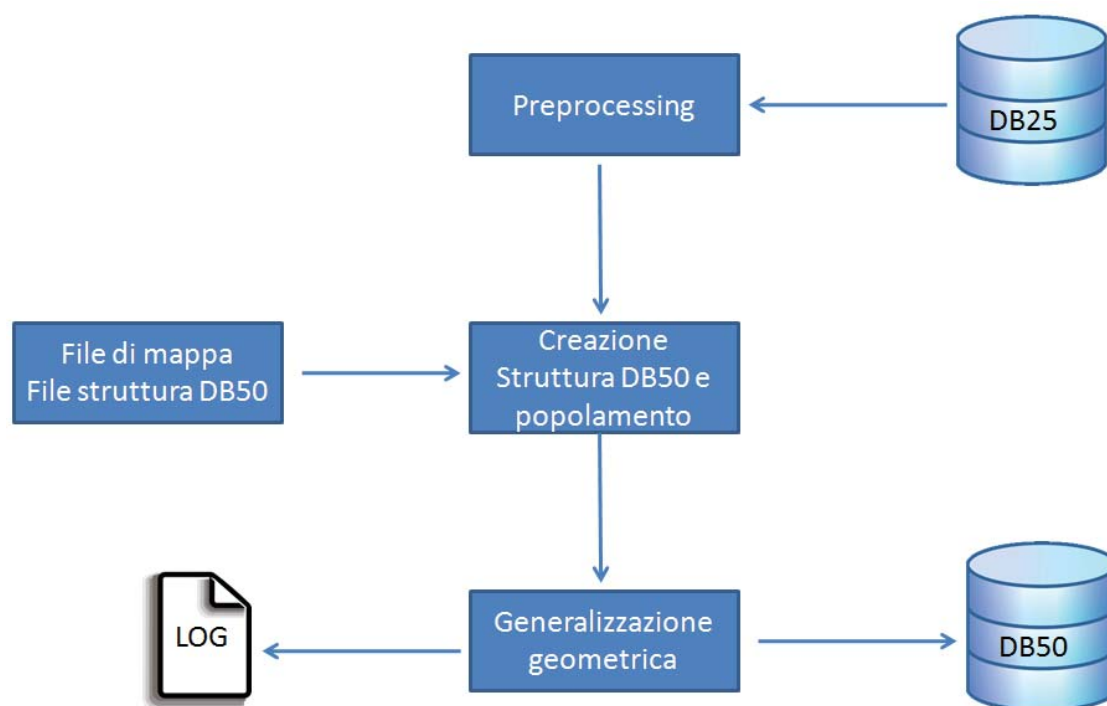


Figura 3.2: Schema del processo

modo più dettagliato nel capitolo successivo, contiene procedure sviluppate ad hoc per elaborare una parte dei dati di input che altrimenti non potrebbero essere correttamente generalizzati se-

manticamente. A questo punto i dati vengono passati al modulo di generalizzazione semantica, che li elabora e li prepara per essere processati dal modulo della generalizzazione geometrica: dopo quest'ultima fase, il risultato è il DB50 popolato con i dati generalizzati per la scala 1.50000. Per la gestione del flusso è stato creato un modulo di esecuzione e di controllo che, oltre a eseguire le chiamate agli altri moduli, svolge importanti funzioni. La prima è naturalmente quella di interfacciarsi con i database per la lettura dei dati di input e la memorizzazione dei dati generalizzati nel DB50. Quest'ultimo viene creato al volo: il modulo riceve in input un file contenente la struttura espressa in meta-linguaggio che viene interpretato da un parser, generando le query per la creazione del database. Il software offre un'interessante alternativa all'elaborazione dei dati memorizzati su database: l'utilizzo di alcuni plugin permette l'esecuzione del processo sul software GIS *OpenJump*, tramite il quale i dati vengono memorizzati in tabelle caricate in memoria e alle quali si accede tramite l'utilizzo di connessioni virtuali. In questo modo i dati vengono gestiti come se risiedessero su un vero database. Questa modalità di esecuzione, oltre a migliorare le prestazioni del processo, permette un'immediata visualizzazione dei risultati non solo dell'intero processo ma anche quelli relativi alle varie fasi. Un'altra funzionalità offerta dal software prodotto è quella di poter utilizzare il *logger*: il modulo di controllo si interfaccia a tale servizio, che permette il *backtracking* sulle modifiche delle geometrie. L'utilizzo del logger, unitamente all'esecuzione del processo su *Openjump*, si è rivelato un strumento utile in fase di debug degli algoritmi e del processo in generale. Il software è inoltre predisposto all'utilizzo del *Partitioning*, ovvero la tecnica che permette di dividere lo spazio cartografico in partizioni da elaborare separatamente, possibilmente senza considerare il contenuto informativo proveniente da partizioni vicine (il tutto senza influenzare il risultato finale. Tale funzionalità è stata in parte già sviluppata all'interno del progetto Cargen ma essendo ancora in fase sperimentale non è stata inserita all'interno del processo.

Dopo aver descritto le varie fasi che hanno portato allo sviluppo del processo, nel prossimo capitolo verrà descritta in modo più dettagliato la fase di generalizzazione semantica, sulla quale il presente lavoro di tesi si è maggiormente focalizzato.

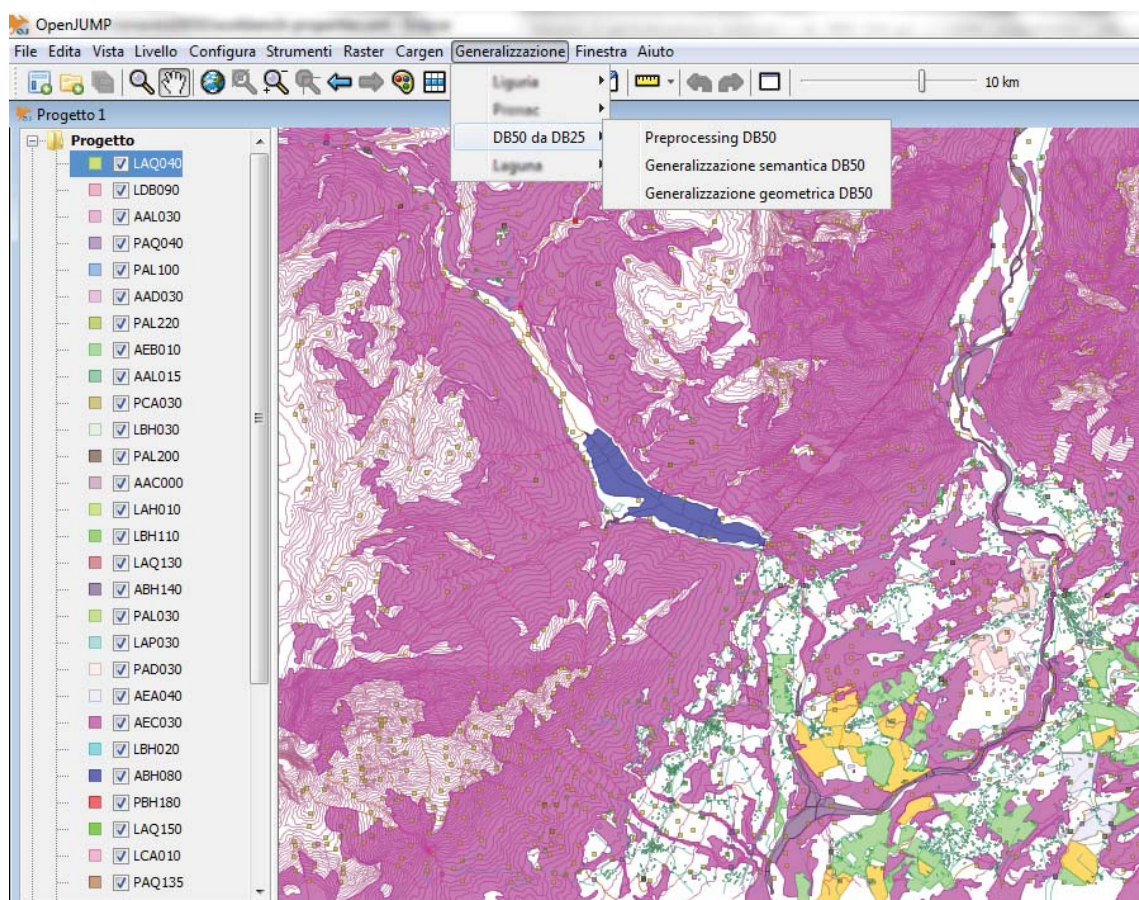


Figura 3.3: Esecuzione del processo su OpenJump. A sinistra, le tabelle del database virtuale

Capitolo 4

Generalizzazione del modello

All'interno dell'intero processo, la generalizzazione del modello traduce il contenuto del database di origine per renderlo coerente con il modello di dati del database di destinazione. Quando si produce una mappa, il cartografo crea un modello astratto della realtà effettuando una selezione dei fenomeni ritenuti rilevanti. Quando si ha a che fare con un database geografico, si parla di modello di dati e definisce gli oggetti della realtà che devono essere presenti nel database (detti classi o *feature classes*) e quali delle loro caratteristiche devono essere descritte (gli attributi svolgono questo compito di descrizione). Due mappe a scale diverse adottano, solitamente, due differenti modelli di dati: più piccola è la scala di una mappa, infatti, e minore è il dettaglio con cui essa rappresenta una determinata realtà. Di conseguenza non tutti i fenomeni rappresentati ad una scala più grande sono rappresentabili in una più piccola, rendendo inevitabile l'utilizzo di due diversi modelli per rappresentare la stessa realtà stessa. Così, un processo di generalizzazione cartografica non solo deve trasformare la rappresentazione grafica degli oggetti della per adattarli alla scala di destinazione, ma anche "tradurre,, le informazioni semantiche secondo il modello di dati di destinazione: questo processo è chiamato generalizzazione semantica. Come è facile intuire, tale processo sarebbe molto semplice se le classi di oggetti nei due modelli fossero in corrispondenza $1:1$; in caso contrario, è necessario effettuare una trasformazione dei dati semantici di origine. Passando ad una scala più piccola è facile che le classi di destinazione siano in corrispondenza $1:n$ con quelle di origine: questo vuole dire che n classi devono essere generalizzate in un'unica classe. Anche se più raramente, è possibile che una classe debba essere mappata in più classi nel modello di destinazione. In entrambi i casi, la generalizzazione semantica utilizza i valori degli attributi che definiscono la specificità degli oggetti - o qualche altra forma di vincoli - per decidere come tradurre i dati del modello di partenza. Può verificarsi anche il caso in cui una classe nel modello di partenza non abbia nessuna corrispondenza nel

nuovo modello, per esempio quando un determinato oggetto non è ritenuto rilevante alla scala della mappa che si vuole derivare; viceversa, una classe del nuovo modello potrebbe non avere una corrispondenza nel modello sorgente quando un fenomeno che si vorrebbe rappresentare non esiste nella scala di partenza. Nel primo caso i dati originali andranno persi e non saranno presenti nei dati generalizzati, nel secondo caso potrebbe essere possibile recuperare le informazioni mancanti da altre classi del modello di partenza o, in caso contrario, c'è un'incompatibilità tra i due modelli e la classe del database derivata non sarà popolata. Oltre a definire le classi degli oggetti e i loro attributi, ciascun modello di dati può contenere dei vincoli sulle caratteristiche degli oggetti. Possono essere specificati vincoli dimensionali (ad esempio, in base ad una certa soglia definita sull'area di un edificio questo può essere rappresentato mediante un punto piuttosto che un poligono), relazioni spaziali o altre condizioni: anche queste specifiche devono essere considerate durante il processo di generalizzazione del modello. Generalmente la compatibilità tra i modelli e, di conseguenza, la complessità della generalizzazione semantica crescono con l'aumentare della differenza tra la scala di origine e quella di destinazione. Nel nostro caso i due modelli di dati, il DB25 e il DB50 mostrano un buon grado di compatibilità e tutte le classi sono derivabili; nonostante le somiglianze, però, è stato necessario lo sviluppo di alcuni algoritmi per derivare le classi del DB50 dai dati di origine. La messa a punto del processo di generalizzazione del modello è avvenuta in due fasi: prima è stata realizzata una mappa tra i due modelli, in seguito sono state create le regole per la derivazione. Entrambe le operazioni sono state effettuate manualmente e il risultato finale è un'applicazione, scritta in linguaggio Java, che copia i dati dal DB25 nel DB50, applicando le trasformazioni necessarie. In questo capitolo verranno descritti i modelli dei dati utilizzati nel processo e i passi che hanno portato all'implementazione del processo di generalizzazione semantica.

4.1 I modelli dei dati

4.1.1 Il DBT

Dal 1977, in Italia, la produzione di carte topografiche a grande scala (1:1000 ÷ 1:10.000) era affidata ufficialmente Regioni. Esse avevano, dunque, la totale autonomia nella scelta delle caratteristiche e della produzione delle mappe topografiche: ogni Regione poteva strutturare i dati cartografici secondo un proprio modello di dati, adottare una propria classificazione degli elementi topografici e decidere quali elementi inserire nel modello e come rappresentarli. Questa situazione ha causato, di fatto, gravi problemi a livello di interoperabilità tra le mappe topogra-

fiche, ognuna semanticamente diversa dalle altre. Nel febbraio del 2012 il Governo italiano, con un decreto di legge denominato DM10/11/2011, ha gettato le basi per l'uniformità, a livello nazionale, della produzione dei database topografici (DBT). Il decreto esprime, infatti, i criteri generali per l'istituzione e mantenimento dei database geografici di interesse generale e definisce, tra le altre cose, un modello dei dati unico per la produzione della cartografia a grande scala. Il modello dei dati del DBT ha una struttura gerarchica, organizzata in Strati, Temi e Classi. Gli strati sono i seguenti:

- Informazioni geodetiche e fotogrammetriche
- Viabilità, mobilità e trasporti
- Immobili ed antropizzazioni
- Gestione viabilità e indirizzi
- Idrografia
- Orografia
- Vegetazione
- Reti di sottoservizi
- Località significative e scritte cartografiche
- Ambiti amministrativi
- Aree di pertinenza

In figura 4.1 la classificazione dello strato "Immobili ed antropizzazioni".

Anche la codifica delle classi, dei loro attributi e dei domini degli attributi ha una struttura gerarchica: nel modello ad ogni classe è attribuito un codice di 6 cifre, ai suoi attributi è attribuito un codice di 8 cifre formato dalle 6 cifre della classe più altre due cifre, e a i valori degli attributi sono codificati con 10 un codice di 10 cifre, le 8 del codice dell'attributo più altre due cifre. Il modello contiene, inoltre, i limiti di acquisizione per le classi a cui devono essere applicati.

```

STRATO: IMMOBILI ED ANTROPIZZAZIONI
TEMA: EDIFICATO
CLASSE: UNITA' VOLUMETRICA
CLASSE: EDIFICIO
CLASSE: CASSONE EDILIZIO
CLASSE: ELEMENTO DI COPERTURA
TEMA: MANUFATTI
CLASSE: MANUFATTO EDILIZIO
CLASSE: SOSTEGNO A TRALICCIO
CLASSE: PALO
CLASSE: ELEMENTO DIVISORIO
CLASSE: MURO O DIVISIONE IN SPESSORE
CLASSE: CONDUTTURA
CLASSE: LOCALIZZAZIONE DI MANUFATTO EDILIZIO O DI ARREDO/IGIENE URBANA
CLASSE: LOCALIZZAZIONE DI MANUFATTO DI RETE TECNOLOGICA
CLASSE: LOCALIZZAZIONE DI MANUFATTO INDUSTRIALE/DI TRASPORTO
TEMA: OPERE DELLE INFRASTRUTTURE DI TRASPORTO
CLASSE: PONTE/VIADOTTO/CAVALCAVIA
CLASSE: GALLERIA
TEMA: OPERE DI SOSTEGNO E DI DIFESA DEL SUOLO
CLASSE: MURO DI SOSTEGNO E RITENUTA DEL TERRENO
TEMA: OPERE IDRAULICHE, DI DIFESA E DI REGIMAZIONE IDRAULICA
CLASSE: DIGA
CLASSE: ARGINI
CLASSE: OPERE IDRAULICHE DI REGOLAZIONE
CLASSE: ATTREZZATURE PER LA NAVIGAZIONE
CLASSE: OPERE PORTUALI E DI DIFESA DELLE COSTE

```

Figura 4.1: Esempio di classificazione nel modello del DBT.

4.1.2 Il DB25

Il modello di dati DB25 è il modello ufficiale IGM per la scala 1:25000, progettato dall'IGM per le nuove mappe della serie DB25 e pensato per essere derivabile dalla cartografia regionale. Il modello non è costruito per la produzione diretta delle mappe della serie, ma in realtà rappresenta un DLM - Digital Landscape Model - dal quale, dopo un'ulteriore elaborazione, possono essere prodotti i DCM - Digital Cartography Model e le mappe: per questo motivo cerca di conservare quanto più possibile la precisione della cartografia regionale. La maggior parte degli oggetti sono rappresentati con elementi puntuali o lineari e solo pochi di essi hanno una geometria poligonale, per lo più edifici e elementi naturali che si estendono su grandi superfici come laghi, grandi fiumi, rocce, boschi etc. Le reti stradali, ferroviarie e fluviali sono rappresentate solo mediante gli archi e nessuna è esplicitamente espressa mediante grafo. Quasi tutti gli oggetti hanno un limite di acquisizione, cioè una soglia minima, che determina se un oggetto deve essere incluso nel database o meno a seconda delle sue dimensioni; i vincoli dimensionali sono utilizzati anche per decidere di che tipo deve essere la geometria che descrive un determinato oggetto, così che due elementi che nel mondo reale possono essere classificati nello stesso modo, nel DB25 potrebbero appartenere a due classi diverse (ad esempio una tettoia può essere classificata come "Tettoia grande" o "Tettoia piccola" in base alla misura della sua superficie e le due classi sono rappresentate rispettivamente da un poligono e da un punto). Il modello ha in totale 149 classi, ognuna delle quali ha un nome e un codice composto da una lettera che indica il tipo di geometria (A: area, L: linea, P: punto) e il codice FACC¹. All'interno di ciascuna classe il modello contiene uno o più oggetti, detti *Label*, ognuno dei quali è una particolare istanza di una classe e

¹Feature Attribute Coding Catalogue, un sistema globale di codifica delle feature, dei loro attributi e dei valori degli attributi realizzato dal DGIWG [DGIWG, 2000]

rappresenta esattamente un tipo di dato della mappa. Ogni oggetto ha un identificatore univoco, memorizzato nell'attributo "LAB", ha una propria definizione e le proprie specifiche. Il numero totale di oggetti distinti è 239.

4.2 Il DB50

A differenza della prima fase del progetto Cargen, che si sviluppava su due modelli di dati definiti - il GeoDBR alla scala 1:5000 e il DB25 IGM alla scala 1:25000 - l'attuale progetto di ricerca non poteva basarsi su alcun modello di dati per la scala generalizzata 1:50000. Questa mancanza ha reso necessaria la progettazione preliminare di un modello dati per un database topografico alla scala 1:50000. Come visto nel capitolo precedente, dopo una prima fase di analisi sono stati fissati i requisiti del modello:

- Possibilità di rappresentare tutti gli oggetti della Carta d'Italia
- Permettere la derivabilità dal DB25
- Classificare gli oggetti secondo la struttura IGM

Per la realizzazione del nuovo modello si è deciso di prendere spunto dalla Carta d'Italia in scala 1:50000 dell'Istituto Geografico Militare e, come primo passo, è stata eseguita un'analisi comparativa tra il DB25 e la Carta d'Italia in scala 1:50000, tracciando una mappatura tra gli elementi topografici contenuti nelle due cartografie, analizzandone la derivabilità, le differenze e le incompatibilità. Nella progettazione del modello dati si è tentato di mantenere distinguibili tutti gli oggetti topografici aventi nella Carta d'Italia in scala 1:50000 un simbolo grafico unico; oggetti simili sono stati raggruppati dentro classi uniche e, all'interno della stessa classe, sono distinguibili tramite attributo. Dopo aver estratto tutti gli oggetti della Carta d'Italia, questi sono stati normalizzati ed è stato quindi possibile formalizzare la struttura in classi. Per conformità con il DB25 si sono mantenuti sia i codici FACC, con la convenzione che ogni codice possa essere usato da più classi a patto che abbiano geometria diversa, sia l'uso dei codici LAB con la stessa accezione usata nel modello del DB25. Una volta individuati gli elementi topografici e classificati secondo la struttura, per ognuno di essi sono stati decisi gli attributi e i loro domini, compreso il tipo di geometria con cui devono essere rappresentati (Polygon, Linestring, Point), e sono stati formalizzati i limiti di acquisizione. Il risultato di tutto queste operazioni è il disegno vero e proprio del modello.

Come è facile aspettarsi, passando ad una scala più piccola e dovendo rappresentare la realtà con minor dettaglio, il modello del DB50 presenta una granularità minore rispetto a quello del

FACC	Nome	Geometria	Attributo	Nome attributo	Formato	n. cifre	Enumerato	Limiti
Strade								
AP030	Road	Line	LAB	Label of Feature	string	6	Enum	lunghezza>200
							L301	Autostrada
							L302	Extraurbana principale
							L303	Extraurbana secondaria
							L304	Strada urbana di scorrimento
							L305	Strada urbana di quartiere
							L306	Strada locale
							L315	Strade secondarie all'interno degli agglomerati urbani
							L318	Autostrada in costruzione
			LOC	Vertical Relative Location	string	3	Enum	
							008	A livello del suolo
							025	Sopra il livello del suolo
							040	Sotto il livello del suolo
							997	Non inserito
			NAM	Name	string	80		
			TXT	Associated Text	string	255		
			WD1	Route Minimum Travelled V	int	*		
			EXS	Existence Category	string	3	Enum	
							028	In esercizio (operativo)
							005	In costruzione
			MED	Median Category	string	3	Enum	
							001	Con spartitraffico
							002	Senza spartitraffico
			RST	Load-bearing Surface Typ	string	3	Enum	
							001	Rivestimento duro
							002	Rivestimento naturale
							003	Rivestimento leggero
			CTF	Classifica Tecnico Funzior	string	3	Enum	
							001	Autostrada
							002	Extraurbana principale
							003	Extraurbana secondaria
							004	Strada urbana di scorrimento
							005	Strada urbana di quartiere
							006	Strada locale

Figura 4.2: Esempio di classificazione nel modello DB50: le strade

DB25. Questo fa sì che alcuni oggetti presenti nel modello di partenza non siano presenti nel nostro modello e che diversi oggetti del DB25 perdano la loro specificità e vengano classificati con lo stesso LAB. Il modello del DB50 contiene 127 classi e 236 *Label* distinti.

4.3 La mappatura

Il processo di mappatura, in cui ci si occupa di trovare le corrispondenze tra le classi del modello sorgente e quello destinazione, richiede generalmente di studiare attentamente le caratteristiche di entrambi i modelli di dati, analizzando le definizioni delle classi, gli attributi ed i valori di questi ultimi.

Chiaramente, per come è stato progettato il DB50, non sono state riscontrate situazioni critiche per quanto riguarda la derivabilità. Alla fine del processo di mappatura, in base al loro grado di derivabilità, sono stati individuati due tipi di oggetti:

- Oggetti derivabili direttamente;
- Oggetti non derivabili direttamente

La maggior parte degli oggetti è, inoltre, soggetta a specifiche.

FACC	Geometria	Attrib	Nome attrib	Formato	n.c.	Enumerato				FACC	Geom	Attributo	Enumerato				
AN010	Line	LAB	Etichetta	text	6	Enum				AN010	Line	LAB	L201	Ferrovie a 1 binario a scartamento ordinario			
						L201	Ferrovie a 1 binario a scartamento ordinario						L201A	Binario all'intero di stazioni o scali ferroviari			
						L202	Ferrovie a 2 o più binari a scartamento ordinario						L202	Ferrovie a 2 o più binari a scartamento ordinario			
						L203	Ferrovie a scartamento ridotto, Tramvia in sede propria,						L205	Ferrovie a scartamento ridotto			
													L207A	Tramvia in sede propria			
													L205A	Funicolare/Cremagliera			
						L204	Tramvia in sede stradale						L207	Tramvia in sede stradale			
						L205	Ferrovie in costruzione						L214	Ferrovie in costruzione			
						L206	Ferrovie in disarmo o in abbandono						L216	Ferrovie in disarmo			
		LOC	Categoria c int		3	Enum						LOC	Enum				
						'008	A livello del suolo						'008	A livello del suolo			
						'025	Sopra il livello del suolo						'025	Sopra il livello del suolo			
						'040	Sotto il livello del suolo						'040	Sotto il livello del suolo			
						'997	Non inserito						'997	Non inserito			
		LTN	Numero di l int									LTN					
		NAM	Nome	text								NAM					
		RRA	Elettrificazi	int	3	Enum		si				RRA	Enum				
						'003	Elettrificata con cavi aerei						'003	Elettrificata con cavi aerei			
						'004	Non elettrificata						'004	Non elettrificata			
						'997	Non inserito						'997	Non inserito			
		TXT	Note	text								TXT					
		EXS	Categoria c int		3	Enum						EXS	Enum				
						'005	In costruzione						'005	In costruzione			
						'008	In disarmo						'008	In disarmo			
						'028	In esercizio (operativo)		si				'028	In esercizio (operativo)			
		RGC	Categoria c int		3	Enum						RGC	Enum				
						'000	Sconosciuto						'000	Sconosciuto			
						'002	Ridotto						'002	Ridotto			
						'005	Standard						'005	Standard			
		RRC	Categoria f int		3	Enum						RRC	Enum				
						'014	Tramvia in sede propria		si				'014	Tramvia in sede propria			
						'015	Tracciato ferroviario inclinato						'015	Tracciato ferroviario inclinato			
						'016	Linea principale						'016	Linea principale			
						'017	Linea secondaria						'017	Linea secondaria			

Figura 4.3: Esempio di mappatura tra i modelli DB25 e DB50: le ferrovie

4.3.1 Oggetti derivabili direttamente

Gli oggetti appartenenti a questo gruppo possono essere derivati utilizzando semplici query in linguaggio SQL, senza ulteriori elaborazioni. Questo significa che, per quanto riguarda questi *Label*, esiste un'ottima corrispondenza tra i due modelli: le loro geometrie sono le stesse ed è necessario, eventualmente, apportare piccole modifiche ai dati semantici (per esempio ai valori di qualche attributo).

4.3.2 Oggetti non derivabili direttamente

Nonostante la grande compatibilità dei due modelli, qualche oggetto del DB50 può non avere nessuna corrispondenza diretta con quelli del DB25. I casi generalmente più semplici sono quelli che legati all'applicazione dei limiti di acquisizione su grandezze facilmente valutabili, come la lunghezza o l'area di una determinata geometria, o all'esecuzione di operazioni che richiedono l'utilizzo di semplici operatori spaziali, come collassamento di un'area nel suo centroide. Ci sono invece casi più complicati, come quando per classificare un oggetto si ha bisogno di analizzare il contesto in cui questo è immerso e come interagisce con gli oggetti ad esso vicini. Per esempio, è possibile voler classificare separatamente (creando quindi due oggetti appartenenti a classi diverse) le tettoie industriali e quelle per uso agricolo. Se nel modello di partenza non sono presenti informazioni sull'uso che viene fatto di quell'oggetto, è possibile intuirne la natura analizzandone il territorio circostante, in particolare la natura degli altri edifici vicini. Questo può

essere fatto creando un *buffer* di una certa grandezza intorno alla tettoia e verificando quindi il tipo di edifici che si trovano ad una determinata distanza da essa. Entrambi i casi hanno richiesto lo sviluppo di codice ad hoc per la generalizzazione semantica dei relativi elementi topografici

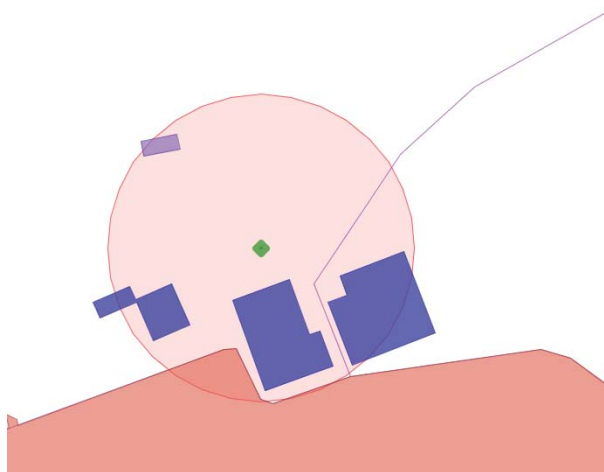


Figura 4.4: Esempio di buffer intorno ad una tettoia per lo studio delle caratteristiche del territorio circostante

4.4 Formalizzazione delle regole e popolamento del DB50

Il passo successivo allo studio dei modelli e alla creazione della mappa è stato quello di popolare le tabelle del DB50 Cargen, sviluppando le procedure che effettuano il trasferimento dei dati dal DB25 al DB50 secondo le regole ricavate nella fase immediatamente precedente. Grazie alle linee guida seguite durante la progettazione del modello DB50 CARGEN, tale fase è risultata nel complesso piuttosto semplice. Il processo sviluppato è completamente automatico e si basa su due moduli sviluppati cercando di produrre una soluzione evoluta e migliorativa di quella utilizzata nella prima fase del progetto Cargen. Le relazioni di derivazione più semplici vengono svolte creando un apposito file di mappa in cui sono formalizzate, con una apposita sintassi, le regole di derivazione semantica; l'utilizzo di questo file permette un più facile inserimento e controllo delle regole di mappatura. Le regole possono essere espresse mediante l'utilizzo di tre funzioni:

- *DEF*: imposta un valore di default per gli attributi. La sintassi per l'utilizzo di questa funzione è la seguente:

```
DEF:DestField1>val1, DestField20>val2
```

Un esempio di utilizzo di questa regola:

```
DEF: nam>tunk, txt>tunk, zv1>d-9999, ao1>d0, loc>t008, zv2>d-9999
```

In questo caso, l'attributo "nam" di tutte le *feature* che lo contengono, se non diversamente precisato in seguito, verrà impostato al valore tunk, l'attributo ao1 sarà impostato al valore d0 etc.

- *SEL*: imposta le regole di selezione. La sintassi ha la seguente struttura:

```
SrcTable:DestTable|FieldName:OldValue>NewValue|SrcField>DestField
```

La prima sezione della riga indica le tabelle (e quindi le classi) del DB25 e del DB50 su cui si sta operando. La seconda indica che il valore dell'attributo FIELD_NAME, che nel DB25 è OLD_VALUE, nel DB50 è invece NEW_VALUE. La terza sezione, infine, indica che il valore dell'attributo SOURCE_FIELD dell'oggetto del DB25 verrà memorizzato nel valore dell'attributo DESTINATION_FIELD dell'oggetto nel nuovo DB.

Una esempio di riga del file di testo è il seguente:

```
LBE015:LBE015 LAB:01:L1008>L1205, L1008B>L1206 crv>crv, txt>txt
```

Questo breve stringa afferma che si sta lavorando sugli oggetti della tabella LBE015 e che, tra questi, quelli identificati nel DB25 dai *LAB* L1008 e L100B saranno identificati rispettivamente con i *LAB* L1025 e L1206 nel DB50. Inoltre gli attributi *crv* e *txt* verranno mappati senza modifiche.

- *TRA* viene utilizzata quando un oggetto, nel passaggio dal modello sorgente a quello di destinazione conserva un certo attributo ma se ne vuole cambiare il valore. La funzione viene richiamata secondo la sintassi:

```
FieldName:OldValue1>NewValue1, OldValue2>NewValue2
```

un esempio:

TRA: HYC:08>01,05>040

Un apposito modulo ha la funzione di interpretare tali regole e tradurle in *query* in linguaggio SQL che possono essere inviate direttamente al server utilizzando la libreria JDBC². Un secondo modulo si occupa invece di eseguire la generalizzazione semantica delle classi che richiedono delle trasformazioni più complesse; in questo modulo sono presenti singoli algoritmi, ognuno sviluppato appositamente per la derivazione di singole classi, che vengono eseguiti in maniera sequenziale. Per alcune classi il processo di derivazione è misto: i dati vengono prima derivati parzialmente applicando una regola scritta nel file di mappa, e successivamente raffinati da un apposito algoritmo. Al termine del processo di generalizzazione semantica i dati si trovano in uno stato ibrido: sono strutturati secondo il modello di riferimento dei dati generalizzati, ma mantengono ancora tutto il contenuto informativo geometrico proprio dei dati a scala maggiore. Il successivo processo di generalizzazione geometrica li renderà conformi alla rappresentazione alla scala minore, completando il processo di generalizzazione.

²Java Database Connectivity, interfaccia per l'esecuzione di istruzioni SQL per le applicazioni JAVA

Capitolo 5

Analisi dei risultati

I test finali, eseguiti su un dataset relativo ad una parte del territorio del Parco delle Dolomiti Bellunesi, hanno permesso di valutare il processo completo per la generalizzazione del DB50 IGM. L'esecuzione del processo sul GIS OpenJump migliora nettamente le prestazioni in termini computazionali, migliorando inoltre la qualità del lavoro di debug. Per quanto riguarda la generalizzazione semantica, il processo si comporta correttamente: l'analisi approfondita dei due modelli dei dati e l'attento studio della derivabilità del DB50 dal DB25 hanno permesso di evitare errori di traduzione delle *feature*. Anche la generalizzazione geometrica ha prodotto dei buoni risultati: gli algoritmi producono dei risultati abbastanza soddisfacenti e, allo stesso tempo, emerge la possibilità di sviluppi e migliorie che contribuirebbero a migliorare la qualità generale del processo. Nelle immagini di seguito vengono riportati alcuni risultati visivi, in cui si mostrano, da sinistra a destra, i dati relativi alla scala 1.5000 e quelli generalizzati alla scala 1:25000 e 1.50000.

Oltre ad evidenziare i buoni risultati del processo, i test hanno fatto emergere alcuni problemi: l'esecuzione del software produce, in alcuni casi, dei risultati non corretti. Alla base di questo comportamento ci sono due principali motivi: errori nei dati di ingresso o condizioni particolari che non sono gestite in maniera ottimale dagli algoritmi di generalizzazione. L'immagine seguente mostra un esempio del primo tipo di problema.

Nell'immagine a sinistra si può notare come la generalizzazione del grafo fluviale abbia disconnesso una parte di esso: questo non è dovuto ad un errore dell'algoritmo ma al fatto che i dati non erano correttamente orientati (la direzione del flusso, dato dalla differenza delle quote dei vertici, era invertito). Nell'immagine a destra si può notare uno *sliver* tra le geometrie di due edifici: questo errore nell'adiacenza delle due forme impedisce al processo di generalizzazione di unire i due edifici come invece è previsto dalle specifiche. Nonostante gli algoritmi adottino



Figura 5.1: *Generalizzazione dell'edificato: in questa serie di tre immagini si possono osservare le trasformazioni subite dall'edificato con il cambio di scala. Nell'immagine centrale si può vedere che i dati originali sono stati semplificati accorpendo edifici adiacenti o limitrofi e semplificandone la forma; solo gli edifici più piccoli sono stati eliminati. Nell'immagine a destra si può notare invece come l'insieme degli edifici sia stato radicalmente trasformato: alcuni edifici sono stati eliminati mentre altri sono stati spostati; gli edifici in genere non sono solo più regolari ma sono anche più grandi ed hanno uno spazio maggiore intorno a loro stessi.*

delle tecniche per poter risolvere situazioni in cui il dato di ingresso presenti delle incongruenze, nei casi in esempio il dato era troppo compromesso per permettere agli algoritmi di rimediare all'errore.

In generale, comunque, i risultati ottenuti possono considerarsi buoni e rappresentano un ulteriore passo in avanti nello studio della generalizzazione automatica e un buon punto di partenza per sviluppi futuri.

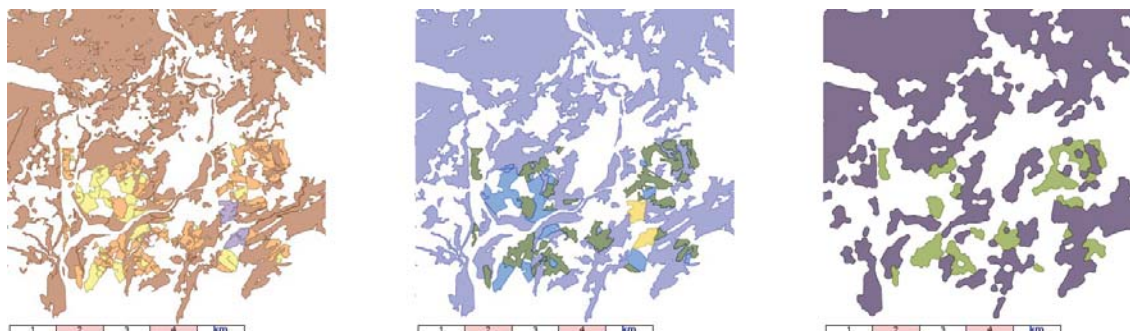


Figura 5.2: Generalizzazione delle grandi aree: in ogni immagine, a diverso colore corrisponde diversa classe. Si può apprezzare come il dato si trasformi scendendo di scala: la frammentarietà del dato si riduce: i poligoni vengono accorpati tra loro od eliminati se troppo piccoli; i buchi interni ai poligoni (ad esempio i broli o le radure dentro un bosco) vengono selezionati in base alla dimensione; la forma stessa dei poligoni viene semplificata, da un lato riducendo il numero di vertici che la compongono, dall'altro eliminando repentini cambi di direzione nel profilo dei poligoni

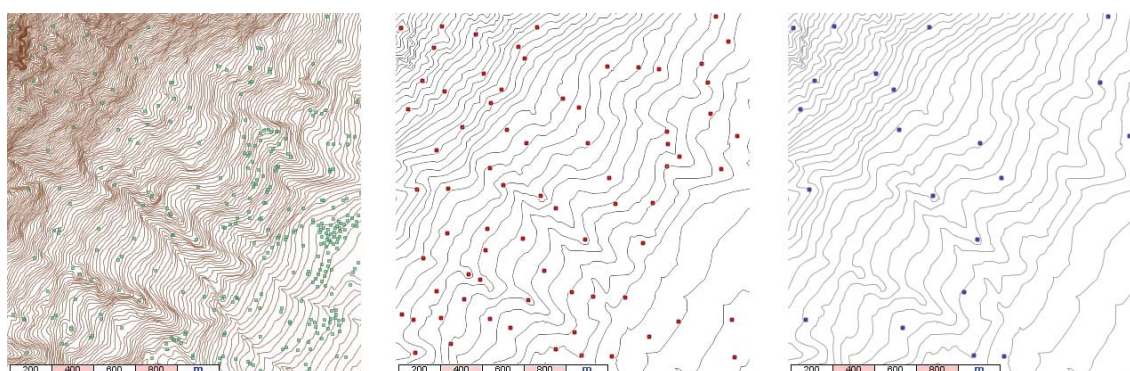


Figura 5.3: Generalizzazione delle curve di livello: le immagini mostrano i punti quotati e le curve di livello. Nelle immagini è possibile osservare la riduzione del numero di punti quotati, sfoltiti sulla base della loro densità ed elevazione, e delle curve di livello, filtrate in base alla quota. Nell'immagine al centro e a destra contengono lo stesso numero di curve poiché i modelli dati rispettivamente del DB25 e del DB50 prevedono entrambi l'acquisizione di curve di livello con quota multipla di 25 metri.

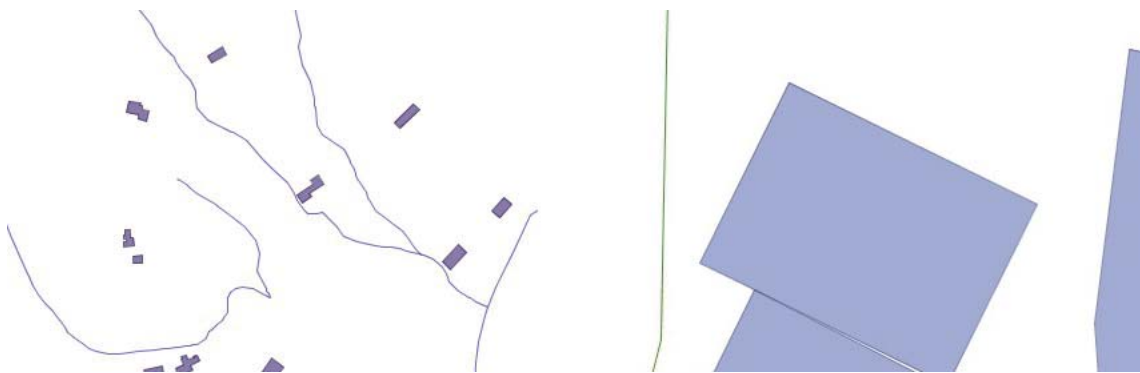


Figura 5.4: Esempio di risultati scorretti della fase di generalizzazione geometrica dovuti a errori nei dati di ingresso

Capitolo 6

Conclusioni e sviluppi futuri

Il lavoro svolto ha fornito una risposta positiva relativamente agli obiettivi prefissati. Le scelte effettuate per quanto riguarda lo sviluppo del modello del DB5 IGM si sono rivelate buone: il modello permette la rappresentazione di tutti gli elementi topografici della carta d'Italia e la completa derivabilità dal modello ufficiale IGM per la scala 1:25000 ha reso più agevole la fase di generalizzazione semantica, permettendo di ridurre a zero gli errori di mappatura degli oggetti. Il processo completo di generalizzazione ha fornito dei buoni risultati, rivelandosi un'ottima evoluzione del processo di generalizzazione del DB25: l'esperienza del lavoro precedente ha permesso di individuare gli aspetti migliorabili di quest'ultimo e di poter intervenire in questo senso, perfezionando dove possibile le funzionalità disponibili e implementandone di nuove. La generalizzazione semantica è resa flessibile dall'utilizzo del *parser* nella fase di mappatura, non obbligando allo sviluppo di nuovo codice in caso di modelli differenti di input e output ma, al contrario, limitando le operazioni alla sola scrittura di nuove regole di mappa. La scelta del modello DB25 come modello di input si è rivelata buona e, avendo a disposizione il processo di generalizzazione del DB25 dal DBT, è possibile studiare in modo più completo i risultati. La modalità di esecuzione sul GIS OpenJump, e non più solo batch, offre la possibilità di una visualizzazione diretta dei risultati del processo, con la possibilità di utilizzare una procedura di vestizione degli elementi della mappa che permette di apprezzarne i dettagli e di verificare ancora più semplicemente e in maniera immediata la correttezza del processo. La struttura a blocchi lo rende il processo flessibile alle modifiche e permette l'esecuzione e la verifica immediata delle singole procedure di generalizzazione: ogni blocco può essere implementato come singolo plugin per OpenJump. L'analisi ha permesso di valutare le prestazioni dell'intero processo e di evidenziare alcuni aspetti migliorabili negli algoritmi di generalizzazione geometrica e di errori nei dati di input: La sensibilità alla qualità di questi ultimi è un problema ancora aperto ed in

fase di studio.

I risultati ottenuti, dunque, non rappresentano una risposta definitiva al problema della generalizzazione automatica ma evidenziano un ulteriore passo in avanti nella tentativo in questo ambito. In futuro si propone il miglioramento degli algoritmi di generalizzazione e l'utilizzo di nuovi dataset riferiti a territori diversi in termini di caratteristiche. Inoltre il processo può essere migliorato con l'integrazione di nuove funzionalità, tra cui le più importanti sono quelle di *Partitioning* e del servizio di controllo automatico della qualità e correttezza dei dati di input. Una maggiore integrazione del processo con il servizio di logging rappresenterebbe un'ulteriore miglioramento della fase di debug degli algoritmi di generalizzazione geometrica.

Appendici

Appendice A

Panoramica degli algoritmi di generalizzazione geometrica

Quella della generalizzazione geometrica è l'ultima fase del processo di generalizzazione. Essa ha lo scopo di ottenere una rappresentazione derivata non solo chiara e leggibile ma coerente con la scala della cartografia che si vuole ottenere. Infatti, la riduzione di scala di una carta è solitamente accompagnata dalla riduzione degli oggetti visibili, così come l'ampliamento e/o l'ingrandimento di oggetti considerati fondamentali che altrimenti non potrebbero essere visibili a quella scala. Come già affermato, il lavoro di tesi non ha richiesto lo sviluppo di algoritmi per la generalizzazione geometrica: essi infatti erano già stati sviluppati nell'ambito del progetto Cargen. Alcuni sono stati sviluppati appositamente per la generalizzazione dei dati alla scala 1:50000, mentre la maggior parte facevano parte anche del processo di generalizzazione del DB25. In quel caso, però, la generalizzazione geometrica nel processo di derivazione del DB25 IGM precedeva la fase di generalizzazione semantica: questo aveva portato allo sviluppo di algoritmi che utilizzavano i dati strutturati secondo il modello del GeoDBR. Nel nuovo processo, gli algoritmi hanno come dati di input quelli già generalizzati semanticamente secondo il modello del DB50. Per poter elaborare dati strutturati secondo un diverso modello, dunque, gli algoritmi sono stati tradotti e adattati inoltre a lavorare per un nuovo obiettivo (una nuova scala). Di seguito verrà data una visione generale degli algoritmi di generalizzazione geometrica, tralasciando i dettagli dell'implementazione, per i quali si rimanda alla documentazione prodotta in fase di sviluppo.

Generalizzazione della rete idrografica Lo strato idrografico, come visto, assume un ruolo chiave nella fase di generalizzazione geometrica e processare questa classe di dati è cercamente

una delle operazioni più complesse dell'intero processo. Questo è dovuto al fatto che, tipicamente, è uno strato contenente un numero elevato di elementi che sono spesso in relazione con quelli di altri strati. A differenza di quanto accadeva per il processo di generalizzazione dei dati in scala 1:25000 dal GeoDBR, i modelli del DB25 e del DB50 classificano gli elementi dello strato nello stesso modo. In base alla larghezza, distinguono i corsi d'acqua in quattro tipi: quelli di larghezza inferiore ad 1 metro, quelli di larghezza compresa tra 1 metro e 5 metri, quelli aventi larghezza tra 5 e 20 metri ed infine quelli più larghi di 20 metri; in entrambi i modelli, i primi tre gruppi sono rappresentati con una linea, l'ultimo mediante una geometria areale. Questo ha fatto sì che il processo di generalizzazione delle rete idrografica non abbia richiesto l'utilizzo di algoritmi per la riclassificazione dei corsi d'acqua. La generalizzazione della rete idrografica è quindi guidata dalle specifiche IGM, che impongono:

- una soglia minima per la lunghezza di fiumi e canali, pari a 300. Tale soglia, come prevedibile, è maggiore rispetto a quella imposta per il modello del DB25.
- lo sfoltimento dei rami della rete nelle zone in cui questi hanno un'elevata densità, che spesso rende difficoltosa la rappresentazione grafica. Lo sfoltimento deve essere applicato favorendo i rami meno importanti tra i candidati all'eliminazione.

Di conseguenza, l'obiettivo della generalizzazione della rete idrografica è quello di applicare tali specifiche, effettuando un *pruning* dei rami fluviali più corti della soglia e di quelli meno importanti nelle zone ad alta densità. La soglia minima per la lunghezza non può essere applicata in modo diretto ai dati perchè l'eliminazione tutti i rami più corti di 300 metri può distruggere il grafo della rete idrografica. E' necessaria dunque una ricostruzione del corso di ogni fiume. Lo sfoltimento nelle zone ad alta densità, invece, richiede una classificazione per importanza dei tratti fluviali. Per fare questo, sono stati sfruttati dei moduli di un'algoritmo sviluppato nella fase precedente del progetto Cargen, che effettuano nell'ordine le seguenti operazioni:

- ricostruzione della direzione del flusso;
- ricostruzione dei tratti fluviali;
- eliminazione dei fiumi di lunghezza inferiore alla soglia minima;
- sfoltimento dei fiumi nelle zone ad alta densità.

Ricostruzione della direzione del flusso Il primo passo per la ricostruzione dei corsi dei fiumi è quello di riconoscere la direzione del flusso. Normalmente questa informazione è inclusa in modo esplicito (come attributo) o implicito (la direzione segue l'ordine dei punti di ogni tratto). Nel caso in cui questa informazione non sia disponibile, deve essere calcolata a partire dalla coordinata z . E questo è il caso del nostro modello di dati. La direzione viene calcolata tramite un algoritmo iterativo che analizza tutti gli archi del grafo fluviale, confrontando la coordinata z dei vertici di ogni arco con quella degli archi a cui esso è connesso. Oltre alla direzione del flusso, con tale procedura, vengono identificate anche le sorgenti, i rami finali dei fiumi e i rami che, per errori dovuti alla precisione della coordinata z , risultano avere il flusso in salita. Tutti questi dati vengono utilizzati dal modulo successivo.

Ricostruzione dei tratti fluviali Utilizzando le informazioni del passo precedente, la ricostruzione dei tratti fluviali inizia con una procedura di arricchimento dei dati, che calcola per ogni tratto di fiume il proprio ordine di Strahler, la distanza dalla sorgente e altri parametri utili per le fasi successive. Combinando questi dati con la classificazione di ogni tratto in base alla larghezza, secondo il modello di dati, ad ogni arco del grafo viene assegnato un punteggio. Partendo dai rami finali, l'idea è quella di trovare ad ogni bivio il ramo più importante del fiume. Una volta giunti alla sorgente, ai tratti percorsi viene attribuito uno stesso codice identificativo (un ulteriore ID). Il meccanismo di punteggio, naturalmente, non è perfetto ma il nostro obiettivo non è quello di ricostruire perfettamente il corso di un fiume, ma farlo con una buona approssimazione.

Eliminazione dei fiumi di lunghezza inferiore alla soglia minima Una volta che il corso di ciascun fiume è stato ricostruito, è semplice individuare quelli di lunghezza inferiore a 300 metri: la lunghezza di ogni fiume viene calcolata come la somma di tutti i tratti a cui è stato attribuito lo stesso codice identificativo nella fase precedente. Se la lunghezza è inferiore alla soglia, tutti i tratti considerati nel calcolo della lunghezza vengono eliminati. Ricorsivamente, ogni fiume derivante da un fiume che è stato eliminato viene eliminato a sua volta.

Sfoltimento dei rami nelle zone ad alta densità L'ultima fase del processo è quello di individuare le regioni dove i fiumi sono troppo vicini gli uni agli altri ed effettuare quindi uno sfoltimento. L'algoritmo utilizza dei buffer per trovare quando i fiumi sono troppo vicini e si basa sui punteggi attribuiti ad ogni fiume nella fase di arricchimento dei dati per calcolare l'importanza di ciascuno di essi. Viene costruito un buffer intorno ad ogni fiume e la percentuale della superficie di questo buffer coperta da altri buffer viene utilizzata per individuare le zone più dense. I fiumi ritenuti poco importanti secondo alcuni parametri e che hanno una percentuale

di buffer coperta sopra una certa soglia vengono eliminati. Ogni volta che viene eliminato un fiume le percentuali vengono ricalcolate e l'algoritmo viene rieseguito fino a quando non ci sono più fiumi candidati all'eliminazione (non ci sono più zone troppo dense o tutti i fiumi hanno un'importanza tale da non poter essere eliminati indipendentemente dalla densità). In figura A.1 i risultati della generalizzazione della rete idrografica.

Generalizzazione della rete stradale La rete stradale è una delle caratteristiche più rilevanti di una mappa e rappresenta probabilmente il segno più riconoscibile di antropizzazione. L'importanza delle strade la loro relazione con altri gli altri elementi di una mappa li rendono uno dei temi principali della generalizzazione. La generalizzazione delle strade affronta in genere con due problemi principali: la selezione e il *displacement*. La selezione permette di ridurre la complessità della rete stradale conservandone un insieme ridotto, il *displacement* permette invece di risolvere i problemi di spazio in modo da poter rappresentare graficamente in modo corretto gli elementi della rete stradale. La generalizzazione delle strade alla scala 1:50000 comprende tre passi principali:

1. semplificazione
2. armonizzazione
3. sfoltimento del grafo stradale

Semplificazione Lo scopo di questa operazione è quello di derivare una strada la cui forma risulti semplice e non ambigua, in modo che rimanga chiara dopo la riduzione di scala. La generalizzazione dell'elemento stradale può avvenire tramite varie azioni:

- *filtraggio*, ovvero la rimozione di alcuni vertici dalla geometria; questa azione è importante anche per ridurre spazio e tempi di calcolo. Per effettuare questa operazione viene utilizzato l'algoritmo di Douglas-Peucker. L'idea alla base dell'algoritmo è abbastanza semplice: per approssimare una linea composta da m vertici, $A_n, A_{n+1}, \dots, A_{n+m}$, l'algoritmo calcola il segmento che congiunge A_n e A_{n+m} e trova tra tutti vertici quello più lontano da quest'ultimo. Se la distanza dal segmento è inferiore a una certa soglia, la linea viene approssimata con il segmento calcolato, altrimenti viene divisa sul vertice più lontano e l'algoritmo ricorsivamente approssima le due parti risultanti.
- *smoothing*, ovvero il *displacement* di alcuni o tutti i punti della linea per attenuarne la forma;

Armonizzazione Per armonizzazione del grafo stradale si intende il processo che porta a risolvere il problema della disomogeneità dei dati. Nel dati di partenza, le strade sono rappresentate mediante archi di un grafo e ad ogni arco è codice che specifica la categoria (strada principale, strada secondaria, strada di campagna, etc.). Può capitare che due archi adiacenti siano classificati in modo diverso: rendere omogenei i dati equivale ad attribuire la medesima categoria ad un gruppo di strade che sono legate da una qualche relazione spaziale o semantica. In figura A.2 un esempio di classificazione errata. L'idea alla base dell'algoritmo di armonizzazione è che la categoria di una strada deve essere costante lungo tutti i tratti che la compongono e che se deve esserci un cambiamento di categoria questo deve avvenire solo in presenza di una condizione particolare (per esempio l'intersezione con una strada di una diversa categoria).

L'idea è stata tradotta in un semplice algoritmo che opera sui tratti costruiti sugli archi del grafo stradale: ogni arco adiacente su entrambi gli estremi a due archi che appartengono alla stessa categoria ma che differiscono dall'arco che si sta considerando, quest'ultimo viene riclassificato e la sua categoria diventerà quella degli archi adiacenti. L'applicazione dell'armonizzazione è limitata da un vincolo: un arco può essere riclassificato solo se più corto di un certo valore di soglia (1000 metri).

Sfoltimento del grafo stradale Secondo quanto specificato dall'IGM, i limiti di acquisizione per le strade alla scala 1:50 000 sono definiti in base alla lunghezza della strada e alla tipologia. Per quanto riguarda la viabilità principale, risultano eliminabili i tratti di lunghezza inferiore a 200 metri. Per la viabilità secondaria, invece, risultano eliminabili le carrarecce di lunghezza inferiore a 400 metri e le mulattiere e i sentieri lunghi meno di 600 metri. La differenza, rispetto alla selezione delle strade relativa alla scala 1:25.000, è che non vengono mantenute le strade *dangling* di lunghezza inferiore ai limiti di acquisizione, anche se sono le uniche a servire un gruppo di edifici. Per evitare di causare una perdita di connessione del grafo stradale, questa regola viene applicata solo alle strade *dangling*, ovvero un segmento stradale che non svolge nessuna funzione di collegamento per altre strade e che ha, quindi, almeno un estremo sconnesso dal resto del grafo. Poiché la rimozione di arco può creare un nuovo arco *dangling*, la regola viene applicata in modo ricorsivo e la ricorsione termina quando la lunghezza dell'arco da eliminare sommata alle lunghezze di tutti gli archi adiacenti già eliminati è superiore alla soglia di 200 metri (questo per evitare che una sequenza di archi *dangling*, ciascuno inferiore a 200 m, causi la rimozione di lunghi tratti stradali).

Generalizzazione della rete ferroviaria Tra le reti di una mappa, quella ferroviaria rappresenta quella meno complessa: sono di gran lunga meno numerosi di fiumi e strade, il loro percorso è

abbastanza regolare e la rete è costituita da sezioni lunghe e dritte con poche intersezioni e cambi di direzione. Il problema principale che bisogna affrontare è la presenza di binari ravvicinati nelle stazioni, che essendo spesso molto fitti ne rendono problematica la rappresentazione. La generalizzazione della rete ferroviaria, dunque, è un tipico problema di tipizzazione: è necessario tracciare una nuova rappresentazione dei binari che sia semplice (cioè che comprenda meno oggetti) ma allo stesso tempo che trasmetta la stessa quantità di informazioni (cioè il le linee generalizzate devono dare ancora l'idea di una stazione). Questo problema non è stato ancora completamente risolto pienamente. L'algoritmo sviluppato cerca di affrontare il problema collassando iterativamente coppie di linee in una linea singola, ma i test hanno evidenziato che non riesce a gestire input complessi e attualmente si sta studiando una nuova soluzione al problema.

Generalizzazione degli edifici Le specifiche IGM per gli edifici impongono uno sfoltimento quando questi sono raggruppati e non sufficientemente distanziati per poter essere rappresentati nel disegno in scala. Gli edifici isolati invece devono essere rappresentati tutti e, se troppo piccoli per essere rappresentati alla nuova scala, ingranditi fino alle dimensioni grafiche minime.

Per la generalizzazione di questo strato informativo è stato utilizzato appositamente per la generalizzazione alla scala 1:50000. Si basa sull'analisi del contesto in cui ogni edificio è immerso. L'algoritmo, servendosi delle reti stradali, fluviale e ferroviarie, crea una partizione dell'intero territorio al fine di poter eseguire un'analisi specifica su una porzione ristretta del territorio. Ogni elemento della partizione, detta faccia, viene classificata come zona urbana ad alta densità, zona urbana a densità media o zona rurale. I tre casi vengono affrontati diversamente. La densità è intesa come rapporto tra la superficie totale della faccia e quella coperta da edifici in essa inclusi. Nelle zone ad alta densità tutti gli edifici vengono uniti a comporre un unico poligono. Questa scelta deriva dal fatto che l'alta densità non permette di effettuare altre operazioni, come il *displacement*. Le zone a densità media vengono invece trattate diversamente: viene prima di tutto effettuata la differenza tra il poligono della faccia e le geometrie degli edifici al suo interno, ottenendo i poligoni che rappresentano gli spazi della faccia non occupati da edifici. Di questi, vengono eliminati quelli con area inferiore ad una certa soglia, mentre gli altri vengono processati dall'algoritmo di Sester per la semplificazione delle geometrie e viene applicata l'operazione di *squaring* per renderli più squadrati. Dopo averle ulteriormente semplificate con altre elaborazioni, le geometrie ottenute rappresentano il nuovo spazio libero all'interno della faccia, il cui resto viene riempito da un poligono rappresentante l'area occupata da edifici. Nelle zone rurali, infine, è possibile applicare l'algoritmo di tipificazione: all'interno di una faccia, gli edifici vengono divisi ulteriormente in gruppi possibilmente omogenei per grandezza e orientamento degli

edifici. Ogni gruppo, in base alle caratteristiche, viene poi processato da una diversa procedura di tipificazione. Di seguito sono descritte brevemente le operazioni coinvolte nei tre metodi di generalizzazione appena descritti.

Selezione Secondo le specifiche IGM, nella generalizzazione dalla scala 1:25000 a quella a 1:50000, la maggior parte degli edifici devono essere mantenuti e che solo quelli non isolati che hanno una un'area più piccola di un certo valore di soglia devono essere cancellati. Per rilevare l'isolamento degli edifici, l'algoritmo disegna un buffer attorno ogni edificio piccolo (cioè con un'area più piccola rispetto alla soglia) e rileva se il buffer contiene altri edifici: se questo si verifica, la struttura non è isolata. Qualsiasi edificio non isolato e con area di sotto della taglia minima edificio viene eliminato.

Fusione Questa operazione è molto semplice, tutti gli edifici adiacenti vengono uniti. Perché questo avvenga devono verificarsi anche altre due condizioni:

- l'intersezione tra due edifici adiacenti deve essere abbastanza grande;
- gli edifici devono essere dello stesso tipo, cioè classificati nella stessa classe secondo il modello del DB50. In figura A.3 un esempio di fusione di edifici adiacenti.

Aggregazione L'algoritmo di aggregazione viene utilizzato per unire due edifici che non sono adiacenti ma non abbastanza distanti da permettere di essere rappresentati separatamente alla nuova scala. Per effettuare l'aggregazione, l'algoritmo lavora come segue:

- (a) viene creato un buffer di raggio R è attorno a ciascuna edificio, dove R è la distanza minima stabilita dalle specifiche IGM. Le intersezioni tra buffer permettono di rilevare vicinanze troppo strette tra edifici ;
- (b) per ogni coppia di edifici candidati all'aggregazione viene calcolata l'intersezione tra i rispettivi buffer;
- (c) viene calcolato l'involuppo convesso dei punti di intersezione;
- (d) si calcolare l'area dell'involuppo convesso: se è troppo piccola, gli edifici non vengono uniti

(e) viene costruito l'MBR¹ dell'involuppo convesso, dalla cui fusione con i due edifici si ottiene la nuova geometria.

Semplificazione La semplificazione degli edifici è richiesta dalle specifiche IGM, che impongono l'eliminazione di rientranze e sporgenze di dimensioni contenute e pertanto non rappresentabili in modo chiaro alla scala considerata. La strategia adottata per effettuare si basa su due algoritmi: uno riduce il numero di punti in ogni geometria, l'altro ne rimuove i piccoli dettagli.

Riduzione dei vertici Per effettuare la riduzione dei vertici viene applicato l'algoritmo di Douglas-Peucker

Eliminazione dettagli La strategia si basa sull'algoritmo ideato da Monika Sester, estendendolo con alcune funzionalità che meglio lo adattano al tipo di dati che si intendono processare. L'algoritmo originale verifica la lunghezza delle facciate di un edificio, definendo una lunghezza critica al di sotto della quale la facciata non risulta più caratterizzante e quindi passibile di rimpiazzo o eliminazione. Le azioni di adeguamento vengono svolte consecutivamente e applicate alla geometria originale di volta in volta. L'algoritmo lavora localmente in accordo con alcune semplici regole, dipendenti dalla geometria dei segmenti adiacenti a quello in esame. La decisione di sostituire o eliminare il segmento in esame dipende dalla geometria che esso forma con il suo precedente e successivo:

- **Inclusione/Esclusione:** l'angolo tra il segmento precedente e quello successivo è di circa 180°. L'azione prevista in questo caso è che il segmento più breve tra il precedente e il successivo venga sottratto quello più lungo, di fatto eliminandolo assieme a quello in esame.
- **Offset:** l'angolo tra il segmento precedente e quello successivo è di circa 0°. L'azione intrapresa è simile al caso di 180°, il segmento più lungo tra il precedente e il successivo viene esteso e di conseguenza il più corto rimosso.
- **Corner:** l'angolo tra il segmento precedente e quello successivo è di circa 90°. Vengono prolungati il segmento precedente e quello successivo fino al punto di intersezione. Anche in questo caso il segmento in esame viene eliminato.

L'algoritmo è stato esteso per permettere la trattabilità di casi in cui le geometrie hanno caratteristiche anche molto diverse da quelle previste nei casi appena discussi. Un esempio di rimozione dei dettagli è illustrato in figura A.5

Squaring L'operazione di *squaring* mira a dare all'edificio un aspetto squadrato, contribuendo a migliorarne l'aspetto estetico senza comunque modificare troppo o snaturare eccessivamente la forma originale. Le operazioni che lo *squaring* esegue sono le seguenti:

- approssimare il più possibile gli angoli della geometria a valori retti (o multipli di esso) o comunque a valori multipli di un angolo scelto inizialmente previa una piccola tolleranza;
- diminuire il numero di angoli esistenti negli edifici.

Come è facile intuire, non tutte le geometrie sono trattabili con questa strategia: gli edifici che hanno forma circolare, non vengono elaborati dal processo, al contrario di quelli con forma più o meno rettangolare. Inoltre ci sono casi in cui applicare lo *squaring* è addirittura sconsigliabile applicarla in quanto snaturerebbe troppo la forma originale della geometria dell'edificio: viene così eseguito preliminarmente un test di applicabilità a tutti gli edifici. In sostanza l'algoritmo è così realizzato, per ogni edificio che supera il test di applicabilità:

- la geometria viene ruotata secondo la sua direzione principale;
- per ogni segmento della geometria viene calcolato il baricentro, la sua lunghezza e l'angolo di rotazione. Quest'ultimo verrà approssimato poi ad un multiplo dell'angolo scelto all'inizio;
- il segmento viene ruotato attorno al proprio baricentro dell'angolo approssimato calcolato al passo precedente;
- vengono modificati i vertici del segmento per rendere la geometria più squadrata.

I vertici degli edifici al termine dell'algoritmo non distano molto da quelli delle geometrie originali, tanto che anche il ridimensionamento delle lunghezze dei lati che si ottiene non è troppo accentuata. Questo permette di effettuare una modifica delle geometrie non troppo invasiva, e quindi di evitare situazioni indesiderate come la sovrapposizione con altre *feature* della mappa.

¹Minimum Bounding Rectangle

Tipificazione L'operazione di tipificazione è un operatore di selezione che cerca di mantenere i modelli spaziali. L'algoritmo cerca di trovare se un edificio è isolato o in un gruppo e se il gruppo ha è distribuito secondo un certo schema: in questo caso si tenta di cancellare alcuni degli edifici e al tempo stesso conservare le informazioni circa la distribuzione spaziale. Per rilevare se un edificio è isolato o in un gruppo, viene utilizzata la stessa procedura applicata nella prima fase di selezione. Quando viene identificato un gruppo di edifici, questo viene processato dall'algoritmo di tipificazione, che può riconoscere modelli lineari e a griglia, con quest'ultimo un caso esteso del precedente. L'obiettivo dell'algoritmo è quello di eliminare alcuni degli edifici di ciascun gruppo in modo da liberare spazio quando un edificio non è abbastanza distante dagli altri. Il primo passo dell'algoritmo è di verificare se esiste una linea che passa attraverso tutti gli edifici del gruppo: il baricentro di ciascun edificio è collegato al baricentro dell'edificio più vicino da un segmento e viene calcolata la media della direzione di questi segmenti, che viene utilizzata per tracciare una linea su ciascuna baricentro del gruppo: se almeno una delle linee attraversa tutti gli edifici, questi sono disposti secondo un modello lineare, altrimenti non lo sono e al gruppo non viene applicata la tipificazione. In caso di disposizione lineare, l'algoritmo iterativo tenta di collocare gli edifici in modo che siano equidistanti tra loro, eliminando eventualmente un edificio ad ogni iterazione se lo spazio non è sufficiente. In caso di disposizione a griglia, gli edifici sono allineati lungo due direzioni principali, generalmente ortogonali tra loro, e ogni edificio appartiene esattamente a due modelli lineari. Per applicare la tipificazione a gruppi di questo tipo, l'idea è quella di considerare il modello come un'estensione 2D di un unico modello lineare: l'algoritmo iterativo applica la generalizzazione ai modelli lineari prima lungo una direzione e poi lungo la direzione opposta. Il processo viene iterato finché la tipificazione porta ad una riduzione del numero di edifici, lungo entrambe le direzioni, che lascia abbastanza spazio libero tra un edificio e quelli ad esso vicini.

Generalizzazione dei fossi I fossi possono distinguersi principalmente in due tipi: quelli isolati e quelli che fanno parte di un gruppo. I primi non vengono processati dall'algoritmo di generalizzazione, mentre agli altri viene applicata la tipificazione. L'algoritmo per prima cosa individua i fossi organizzati in gruppi e, all'interno di ogni gruppo, valuta la vicinanza tra i fossi e le loro direzioni. Se i fossi si trovano ad una distanza inferiore ad una certa soglia e le direzioni sono abbastanza simili, il gruppo viene tipificato. L'algoritmo di tipificazione utilizzato prevede la creazione di nuove geometrie che sostituiranno quelle vecchie, pur conservandone con qualche approssimazione la direzione. Questo è ciò che lo diversifica dall'algoritmo di selezione.

Generalizzazione delle grandi aree irregolari Le grandi aree, nel modello del DB50, rappresentano generalmente boschi, campi coltivati, cave, laghi, ghiacciai ed altre feature. Le operazioni coinvolte nel processo di generalizzazione di queste feature sono le seguenti:

- semplificazione;
- aggregazione e selezione;
- estensione dei confini delle aree;
- collassamento di un'area in una linea

non tutte vengono applicate alla stessa classe di oggetti: la semplificazione viene applicata sempre, mentre l'aggregazione e l'estensione dei confini lineari escludono il collassamento a linea. La semplificazione, che permette di eliminare dalla geometria i punti non necessari alla rappresentazione, si ottiene applicando l'algoritmo di Douglas-Peucker. L'aggregazione delle aree vicine viene effettuata al fine di ridurre il numero di aree da processare, eliminare gli spazi liberi circondati da grandi aree e lo smussamento dei contorni di queste ultime. Questa operazione viene effettuata con l'uso dei buffer: ogni area viene ingrandita creandogli un buffer intorno e le aree che a questo punto si sovrappongono vengono unite. A questo punto viene applicato un buffer negativo alla nuova geometria in modo da riportare le aree alla loro grandezza iniziale ma adesso unite. Infine, viene effettuata la selezione delle aree, eliminando quelle inferiori ad una certa soglia imposta dalle specifiche IGM. Le geometrie finali sono semplificate con una nuova applicazione dell'algoritmo di Douglas-Peucker. L'estensione dei confini è imposta dalle specifiche IGM, che impongono che quando la distanza tra i confini di due grandi aree è sotto una certa soglia, questi devono essere spostati per essere coincidenti. L'idea dell'algoritmo è quella di creare una nuova geometria tra le due di partenza, che poi verrà unita ad una delle due facendo risultare queste ultime adiacenti. L'algoritmo utilizza la Triangolazione di Delaunay Vincolata (CDT): questa viene creata utilizzando i punti dei confini delle due aree e il risultato è un insieme di triangoli contenuti nello spazio tra le due aree. A questo punto, ogni triangolo che ha un lato o un vertice su uno dei confini viene aggiunto alla nuova geometria. Anche l'operazione di collassamento di un'area in una linea viene eseguita ricorrendo alla Triangolazione di Delaunay Vincolata: questa viene effettuata utilizzando i vertici del poligono che rappresenta l'area interessata. Il risultato è un insieme di triangoli con lati sui confini del poligono (lati vincolati) e lati contenuti nel poligono (lati svincolati). Per ogni triangolo, viene tracciati dei segmenti che uniscono ogni lato svincolato al baricentro del triangolo o al vertice opposto al lato, a seconda di alcune caratteristiche del triangolo. L'unione di questi segmenti produce una linea che, dopo

essere stata semplificata mediante una versione modificata dell'algoritmo di Douglas-Peucker, rappresenterà la nuova geometria che sostituisce quella originale.

Generalizzazione degli elementi lineari La generalizzazione di questi elementi è mirata principalmente alla loro semplificazione. In caso di linee parallele ravvicinate, un semplice algoritmo le collassa in una sol linea. Per quanto riguarda le curve di livello, in base alle specifiche IGMI, la loro generalizzazione è un processo semplice che richiede solo una selezione sulla base della loro altezza (questa viene eseguita nella fase di popolamento del database di destinazione recuperando l'informazione dell'altezza da un apposito attributo della feature. Tuttavia si è scelto di semplificare le curve di livello, riducendo il numero di vertici che le compongono, ricorrendo ancora una volta all'algoritmo di Douglas-Peucker. In questo modo viene ridotto notevolmente il numero di dati da processare quando si ha a che fare con le curve di livello.

Generalizzazione degli elementi puntuali Questo tipo di generalizzazione viene effettuata per ridurre il numero di elementi puntuali di determinate classi, come alberi e quote altimetriche. L'algoritmo è semplice e si basa ancora una volta sulla Triangolazione di Delaunay Vincolata.

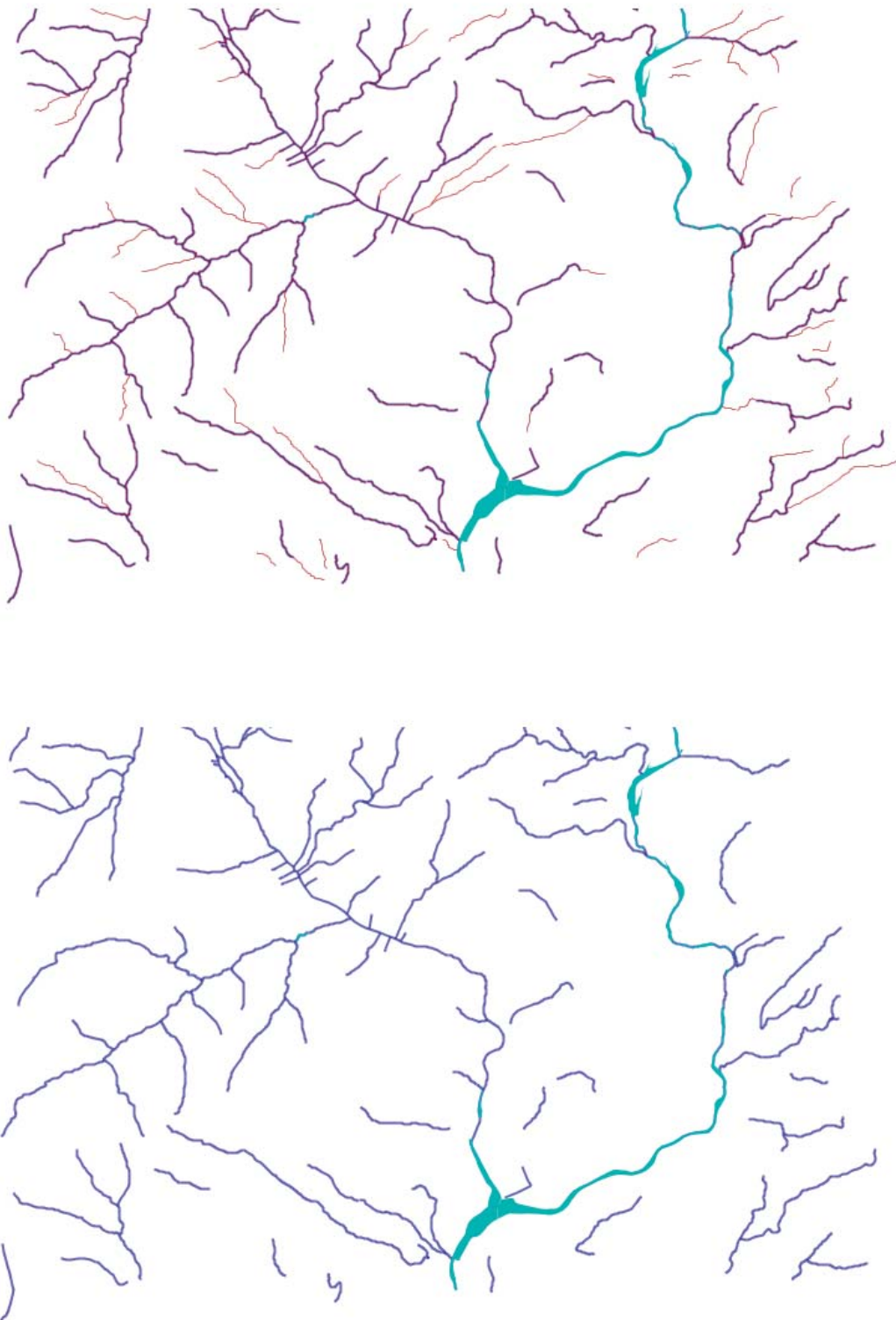


Figura A.1: Nella prima immagine, la rete idrografica alla scala 1:25000: in rosso i rami individuati per il pruning. Nella seconda immagine il risultato della generalizzazione.

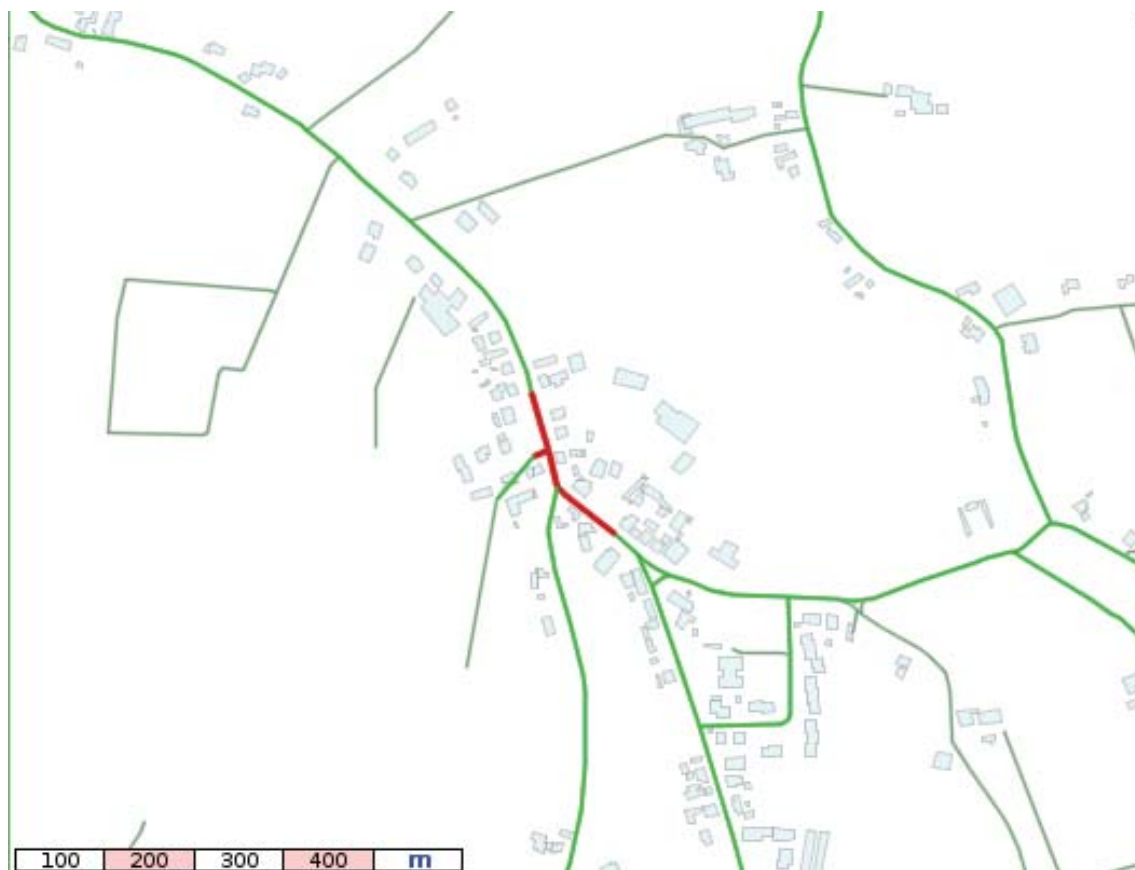


Figura A.2: esempio di strada classificata erroneamente come strada principale (strada rossa).

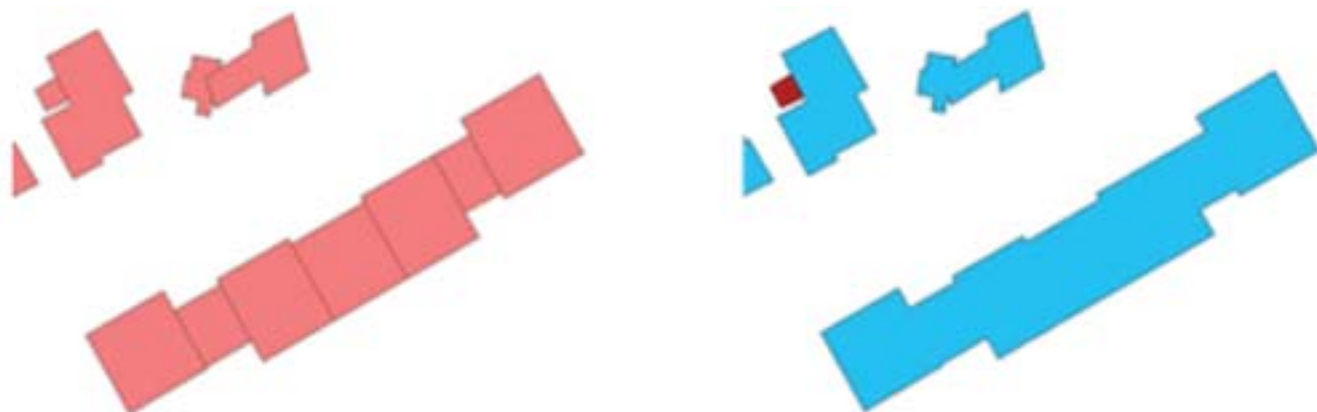


Figura A.3: Fusione di edifici adiacenti. A sinistra i dati di partenza, a destra il risultato della fusione.

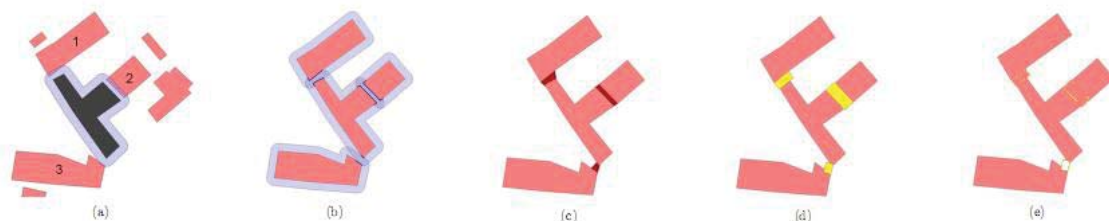


Figura A.4: Esempio di flusso dell'algoritmo di aggregazione di edifici.

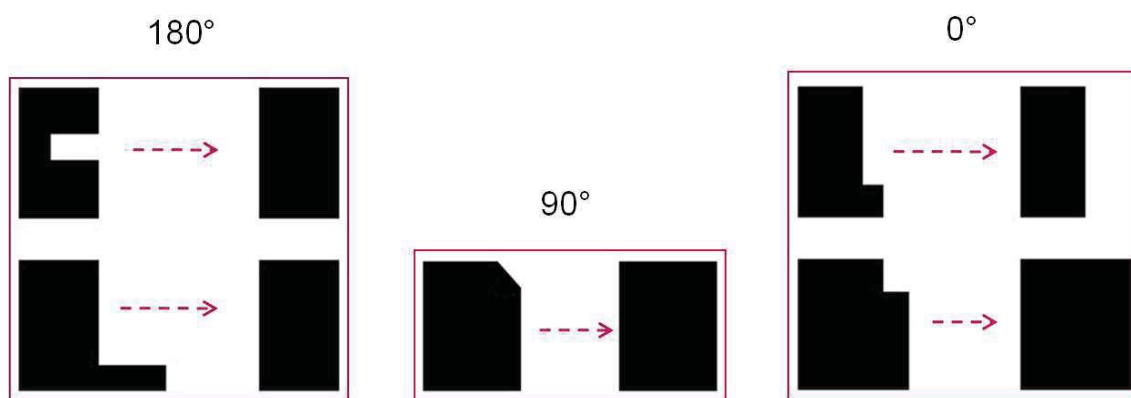


Figura A.5: Il comportamento dell'algoritmo di Sester.

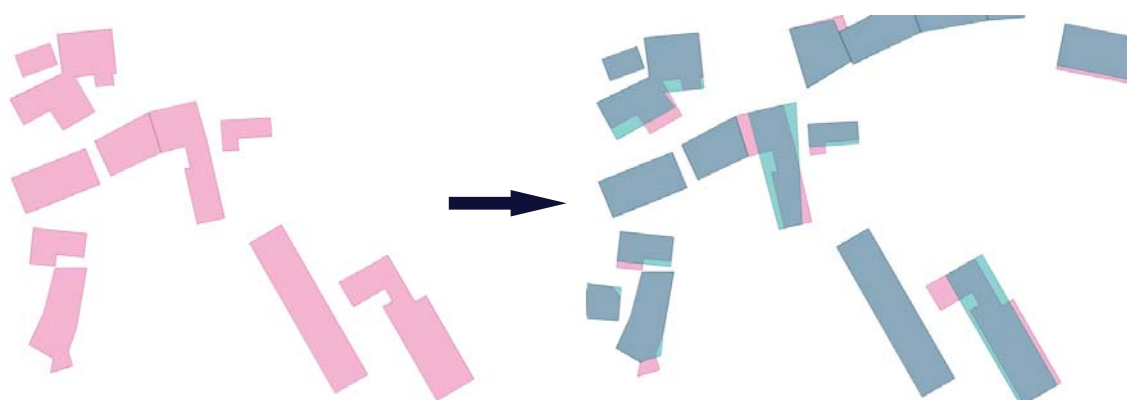


Figura A.6: Risultato della rimozione dei piccoli dettagli nell'algoritmo per la generalizzazione degli edifici

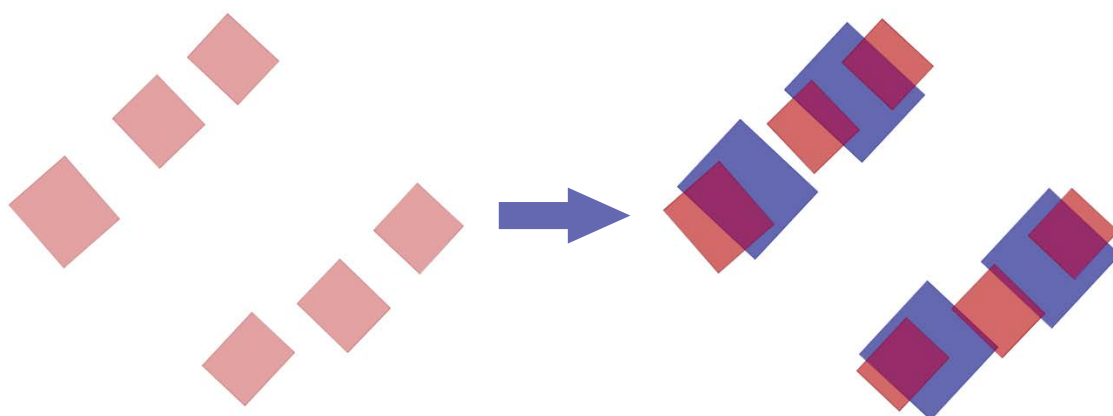


Figura A.7: Risultato della tipificazione di un gruppo di edifici disposti a griglia.

Bibliografia

AGENT (2000) Project ESPRIT AGENT Map Generalisation by Multi-Agent Technology

CARGEN (2009) La generalizzazione del DB25 a partire dal GeoDBR 1:5000 - Schema della derivazione, query per il popolamento e operazioni sulle geometrie, Dipartimento di Ingegneria dell'Informazione, Padova

CNIPA (2009) Centro Nazionale per l'Informatica nella Pubblica Amministrazione Catalogo di dati territoriali – Specifiche di contenuto per i DB Geotopografici

De Gennaro M., Rumor M., Savino S., Le procedure per la derivazione del DB25 dal DBT della Regione del Veneto: risultati del progetto CARGEN, Bollettino della Associazione Italiana di Cartografia, 135, Aprile 2009

DGIWG (2000) The Digital Geographic Information Exchange Standard (DIGEST): Feature and Attribute coding catalogue Digital Geographic Information Working Group

Douglas D., Peucker T. (1973) Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, in Cartographica: The International Journal for Geographic Information and Geovisualization, 10, 112-122

Duchêne C., Ruas A. (2001) Généralisation de données géographiques. Présentation des résultats du projet AGENT Bulletin d'information de l'IGN, n°72 : Bilan de la Recherche

ESRI (1996) Automation of Map Generalization, The cutting-Edge Technology ESRI White Paper Series

ESRI (2004) What is ArcGIS ESRI Press, ISBN 1-58948-090-2

ESRI (2008) ArcObjects JavaDOC

Fécher A., De Waele J. (2006) The future production of generalised maps at IGN Belgium Multiple representation and interoperability of spatial data ISPRS Workshop, Hannover, Germany, 22-24February

Foerster T., Stoter J., Kraak M.-J. (2010) Challenges for automated generalisation at European Mapping Agencies: a qualitative and quantitative analysis The Cartographic Journal, 47 (1), 41-54

GEMURE (2005) The GEMURE project. Multi-scale generalization and multiple representation techniques for efficient production and delivery of cartographic data Final Report ESRI White Paper Series

GeoTools (2002) GeoTools, The Open Source Java GIS Toolkit

Gruenreich D. (1985) Computer assisted generalisation, Papers CERCO-Cartography course, Institut fur Angewandte Geodäsie, Frankfurt a.M.

- Gruenreich D. (1995) Development of computer-assisted generalization on the basis of cartographic model theory, in: Mueller J. C., Lagrange J. P., Weibel R. (Editors), GIS and generalization: methodology and practice, Taylor&Francis, 47-55
- Harrie L., Weibel R. (2007) Modelling the overall process of generalization in: Mackaness W. A., Anne Ruas A., Sarjakoski L. T. (Editors), Generalisation of geographic information: Cartographic modelling and applications, Elsevier Science, 67-87
- Hershberger J., Snoeyink J. (1992) Speeding Up the Douglas-Peucker Line-Simplification Algorithm, in Proceedings of the 5th International Symposium on Spatial Data Handling, 134-143
- IGMI (2006) Struttura del DB25: feature attributi e domini Istituto Geografico Militare Italiano, 2a Direzione della Produzione e Direzione Lavori Ricerca e Sviluppo, Firenze
- Intesa GIS / WG 01 (2004) Specifiche di contenuto: gli strati, i temi e le classi In: specifiche per la realizzazione dei database topografici di interesse generale
- Istituto Geografico Militare (2006) Struttura del DB25: feature attributi e domini Seconda direzione della produzione e direzione lavori di ricerca e sviluppo.
- JTS (2002) Java Topology Suite, Vivid Solutions
- Jabeur, N., Moulin, B. (2004) A multiagent based approach for the resolution of spatial conflicts European Workshop on Multi-Agent Systems (EUMAS), Barcelona, 333-343
- Jahard Y., Lemarié C., Lecordix F. (2003) The implementation of new technology to automate map generalisation and incremental updating processes Proceedings of the 21st International Cartographic Conference, Durban, South Africa, 10-16 August, 1449-1459, CD-ROM
- Joao E. M. (1995) Importance of quantifying the effects of generalization in: Mueller J. C., Lagrange J. P., Weibel R. (Editors), GIS and generalization: methodology and practice, Taylor&Francis, 183-193
- Lecordix F., Le Gallic J. -M., Gondol L. (2007) Development of a new generalisation flow line for topographic maps 10th ICA Workshop on Map Generalization, Moscow, Russia, 2-3 August
- Lecordix F., Lemarié C. (2007) Managing generalisation updates in IGN map production in: Mackaness W. A., Anne Ruas A., Sarjakoski L. T. (Editors), Generalisation of geographic information: Cartographic modelling and applications, Elsevier Science, 285-300
- Mackaness, W.A. (1997) Design as process - strategies for automated map design, Position paper for the 2nd workshop on progress in automated map generalization, Gävle, Sweden.
- Mackaness W. A., Ruas A. (2007) Evaluation of the Map Generalization Process, Generalisation of geographic information: Cartographic modelling and applications, William A. Mackaness, Anne Ruas, L. Tiina Sarjakoski (Editors), Elsevier Science; 1st edition (June 6, 2007), 89-111
- McMaster R. B., Shea K. S. (1988) Cartographic generalization in a Digital Environment: a framework for implementation in a geographic information system GIS/LIS'88 Proceedings, San Antonio, Texas, November 30-December 2, 240-249

- McMaster R. B., Shea K. S. (1992) Generalization in Digital Cartography Association of American Geographers, Washington D.C
- McMaster R. B. (1987) Automated line generalization, Cartographica, 24 (2), 74-111
- McMaster R. B. (2003) The Creation of Multiple Scale Databases in the NHGIS ICA Workshop on Map Generalization, Paris, 28-30 August
- Müller J. C. (1991) Generalisation of Spatial Databases In Geographical Information Systems, D. J. Maguire, M. Goodchild and D. Rhind, (Editors) London: Longman Scientific
- Müller, J. C., Weibel R., Lagrange J. P., Salgé F. (1995) Generalisation: State of the Art and Issues in: Mueller J. C., Lagrange J. P., Weibel R. (Editors), GIS and generalization: methodology and practice, Taylor&Francis, 3-18
- Müller, J. C. (1991) Generalisation of Spatial Databases In Geographical Information Systems, D. J. Maguire, M. Goodchild and D. Rhind, (ed) London: Longman pp. Scientific
- OpenJUMP (2004) Jump Pilot Project online: <http://www.openjump.org/>
- Oracle (2005) Spatial Java API Reference, Online:
http://download.oracle.com/docs/cd/B19306_01/appdev.102/b14373/toc.htm
- Oracle (2005) Spatial User's Guide and Reference online:
http://www.oracle.com/pls/db102/to_pdf?pathname=appdev.102%2Fb14255.pdf&remark=portal+%28Books%29
- Painho M. (1995) Effects of generalization on attribute accuracy in: Mueller J. C., Lagrange J. P., Weibel R. (Editors), GIS and generalization: methodology and practice, Taylor&Francis, 194-206
- Pantazis D., Donnay J.-P, Billen R. (2006) GIS design and implementation methods: The MECOSIG improvement and updating project 25th Urban Data Management Symposium, UDMS, Aalborg, May 15-17
- Refractions Research, (2010) PostGIS JDBC JavaDOC, online:
<http://postgis.refractions.net/documentation/javadoc>
- Regione Veneto (2004) Carta Tecnica Regionale Numerica: codifiche Segreteria Regionale per il Territorio, Unità complessa per il Sistema Informativo Territoriale e la Cartografia
- Regione Veneto (1992) Carta Tecnica Regionale Numerica - Codifiche Regione del Veneto, Segreteria regionale per il territorio, Unità complessa per il Sistema Informativo Territoriale e la Cartografia. online: <http://www.regione.veneto.it/NR/rdonlyres/26AFDADC-13AE-4584-9BA4-3FF395CE2A22/0/Codifiche5000e10000.pdf>
- Regione Veneto (2008) DB Topografico Segreteria Regionale all' Ambiente e Territorio, Unità di Progetto per il SIT e la cartografia. online:
<http://www.regione.veneto.it/NR/rdonlyres/50B13921-AD61-41A0-98CD-9429940E7B38/0/DBTopografico26.pdf>

- Regnauld N., McMaster R. B. (2007) A Synoptic view of generalisation operators in: Mackaness W. A., Anne Ruas A., Sarjakoski L. T. (Editors), Generalisation of geographic information: Cartographic modelling and applications, Elsevier Science, 37-66
- Ruas A, Plazanet C. (1997) Strategies for automated generalization In: Kraak M. J., Molenaar M., Fendel E. M. (Editors), Advances in GIS Research II, Proceedings of the 7th International Symposium on Spatial Data Handling, Delft, The Netherlands, Taylor & Francis, 319-336
- Ruas A. (2000) Project AGENT: Overview and Results of a European R&D Project in Map Generalisation ICA Workshop, Barcelona, Spain, 21-23 November
- Ruas A. (2001) Automatic Generalisation Project: learning process from interactive generalisation, OEEPE, European ORganization for Experimental Photogrammetric Research, Official Publication (39), Frankfurt a.M.:Bundesamt fur Kartographie un Geodaesie
- Ruas, A, Plazanet, C. (1996) Strategies for automated generalization In: Proceedings of the 7th International Symposium on Spatial Data Handling, Advances in GIS Research II, 1996, 6.1-6.18
- Ruas, A., Lagrange, J.P. (1995) Data and knowledge modelling for generalisation, in: Müller J. - C., Lagrange J. -P., Weibel R. (Editors) GIS and generalization, Methodology and Practice, GISDATA 1, Taylor & Francis, 73-90
- Rumor M., Savino S., Congiu S., 2011, The CARGEN project: cartographic generalization as an example of cooperation between National and Regional mapping agencies, UDMS 2011, Delft, The Netherlands (accepted)
- Rumor M., Savino S., Congiu S., De Gennaro M., Zampieri A., 2011, La generalizzazione cartografica automatica: la soluzione per la cartografia del futuro, 150 Anni di Cartografia in Italia, Convegno Nazionale Associazione Italiana Cartografia, Modena, Italia
- Rumor M., Savino S., Congiu S., De Gennaro M., Zampieri A., 2011, Risultati e progressi nella generalizzazione cartografica in Italia: il progetto CARGEN, 150 Anni di Cartografia in Italia, Convegno Nazionale Associazione Italiana Cartografia, Modena, Italia
- Savino S., Rumor M., Congiu S., Tecniche di generalizzazione cartografica dalla grande alla media scala, 15a Conferenza Nazionale ASITA, Parma, Italia (accepted)
- Sester M., Brenner C. (2000) Typification Based On Kohonen Feature Nets Proceedings Of The 1st International Conference On Giscience, Savannah, 21–22
- Shea K. S., McMaster R. B. (1989) Cartographic generalization in a Digital Environment: when and how to generalize AUTO-CARTO, 9th International Symposium on Computer Assisted Cartography, Baltimore, Maryland, April 2-7, 56-67