

UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI INGEGNERIA



Corso di Laurea Magistrale in Ingegneria Informatica

*Un approccio al posizionamento di etichette  
su cartografia generata da database  
cartografici*

***Relatore:***

Prof. Massimo Rumor

***Correlatore:***

Ing. Sandro Savino

***Tesi di:***

Stefano Lazzaro  
Matr. 603622-IF

*Anno accademico 2010-2011*



A mamma, papà, Nicola, Marta,  
per avermi aiutato a  
diventare un uomo.  
A Serena,  
per avermi fatto diventare  
un uomo più dolce :)



## SOMMARIO

LO SCOPO DEL DOCUMENTO è quello di presentare il problema del *Label Placement* nel campo della generalizzazione cartografica, analizzandone la storia, l'evoluzione e le soluzioni, presentando inoltre una soluzione al posizionamento di etichette su cartografia digitale all'interno del progetto CARGEN del Dipartimento di Ingegneria dell'Informazione dell'Università di Padova. Il *Label Placing* è un argomento di grande importanza ed attualità e in questo testo se ne darà una panoramica completa. Nella prima parte del documento sarà quindi introdotto il problema, mentre nella seconda parte è contenuta la presentazione di un possibile approccio risolutivo al problema implementato in Java. Quindi si potranno analizzare i risultati ottenuti, con considerazioni sui possibili sviluppi futuri della soluzione proposta.

# Indice

<b>1. Introduzione</b>	<b>11</b>
<b>2. Cartografia e generalizzazione cartografica</b>	<b>13</b>
2.1 La cartografia . . . . .	13
2.2 Il processo cartografico . . . . .	14
2.3 I GIS (Geographic Information Systems) . . . . .	16
2.4 La generalizzazione cartografica . . . . .	18
2.4.1 Il processo di generalizzazione . . . . .	21
2.4.2 Generalizzazione automatica . . . . .	22
2.4.3 Modelli di generalizzazione . . . . .	23
<b>3. Il progetto CARGEN</b>	<b>26</b>
3.1 La situazione della cartografia in Italia e CARGEN . . . . .	26
3.2 Ambiente di lavoro . . . . .	28
3.2.1 JTS Topology Suite . . . . .	28
3.2.2 OpenJump . . . . .	29
<b>4. Il problema del Label Placement</b>	<b>30</b>
4.1 Introduzione all'Automatic Label Placement e problematiche annesse . . . . .	30
4.2 Panoramica sulla letteratura . . . . .	33
<b>5. Il trattamento della toponomastica nella cartografia digitale</b>	<b>35</b>
5.1 Introduzione al problema . . . . .	35
5.2 Come sono trattati i toponimi nei database cartografici . . . . .	36
5.3 La geometria dei toponimi . . . . .	37
5.4 Forma e contenuto . . . . .	38
5.5 La trattazione con i software commerciali . . . . .	39
5.6 I toponimi nel database secondo IGM . . . . .	40
<b>6. Approccio di posizionamento etichette su cartografia digi- tale in CARGEN</b>	<b>42</b>

---

6.1	Scopi e strumenti di lavoro . . . . .	42
6.2	Le etichette . . . . .	43
6.2.1	Qualità di un'etichetta in base alla posizione rispetto al suo riferimento . . . . .	45
6.2.2	Come si spostano le etichette . . . . .	47
6.3	Valutazione densità di elementi sottostante alle etichette sulla mappa . . . . .	48
6.3.1	La matrice di densità . . . . .	48
6.4	L'algoritmo di piazzamento e il concetto di priorità . . . . .	54
6.4.1	Come vengono individuate le sovrapposizioni tra etichette	54
6.4.2	Le priorità delle etichette . . . . .	54
6.4.3	Generazione matrice di densità e tabella etichette . . .	55
6.4.4	La felicità dell'etichetta singola e quella generale . . .	55
6.4.5	Le ridondanze nelle etichette . . . . .	57
6.4.6	I passi dell'algoritmo . . . . .	58
6.5	Risultati e considerazioni . . . . .	59
6.5.1	Il layer di test . . . . .	59
6.5.2	Interfaccia di test . . . . .	60
6.5.3	Risultati e confronti . . . . .	60
<b>7.</b>	<b>Test su dati reali</b>	<b>64</b>
<b>8.</b>	<b>Conclusioni e possibili sviluppi futuri</b>	<b>69</b>
	<b>Bibliografia</b>	<b>71</b>

# Elenco delle tabelle

# Elenco delle figure

2.1	Esempio di generalizzazione di una mappa . . . . .	19
2.2	Fasi della generalizzazione in un GIS . . . . .	21
4.1	Andamento pubblicazioni su Automatic Label Placement . . .	31
4.2	Problemi di sovrapposizione etichette . . . . .	32
4.3	Posizionamento etichetta su punto su circonferenza . . . . .	33
4.4	Line labeling di Wolf . . . . .	34
5.1	Difficoltà standard toponomastica . . . . .	36
5.2	Possibili stampe di toponimi . . . . .	38
6.1	Come vengono stampate le etichette su OpenJump . . . . .	44
6.2	Posizioni etichetta puntuale . . . . .	45
6.3	Posizioni etichetta areale . . . . .	46
6.4	Posizioni etichetta su linea spezzata . . . . .	46
6.5	Spostamenti dell'etichetta areale . . . . .	47
6.6	Posizioni etichetta su linea spezzata . . . . .	48
6.7	Illeggibilità etichetta dovuta a sovraffollamento sulla mappa. .	49
6.8	Matrice di densità errata . . . . .	50
6.9	Illustrazione funzionamento algoritmo di densificazione . . . .	51
6.10	Matrice di densità efficace . . . . .	51
6.11	Realizzazione della matrice di densità su di un layer . . . . .	52
6.12	Esempio matrice di densità . . . . .	53
6.13	Bordi di un'etichetta per le sovrapposizioni . . . . .	54
6.14	Funzione di calcolo densità vertici sotto etichetta . . . . .	56

---

6.15	Ridondanza etichette . . . . .	57
6.16	Ridondanza etichette risolta . . . . .	58
6.17	Layer di test dell'algoritmo . . . . .	60
6.18	Pannello di controllo algoritmo . . . . .	61
6.19	Risultati dell'algoritmo di posizionamento . . . . .	62
6.20	Risultato grafico secondo passaggio algoritmo. . . . .	63
6.21	Dettaglio risultati algoritmo su porzione di mappa . . . . .	63
7.1	Dati di test. Sezione DB25 di Chioggia . . . . .	65
7.2	Risultati generali su test Chioggia . . . . .	67
7.3	Dettaglio risultati su Chioggia . . . . .	68



## 1. INTRODUZIONE

La realizzazione di una mappa geografica è un processo molto elaborato che si divide in più parti. La fase di generalizzazione riveste un ruolo fondamentale. Lo scopo di questa fase è 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 è necessaria perché, in una mappa, non è generalmente possibile rappresentare tutti i fenomeni geografici: lo spazio per contenerli, infatti, diminuisce con il diminuire della scala.

Spesso, nella produzione di una cartografia a bassa scala, l'astrazione della realtà avviene a partire dalle informazioni geografiche contenute in una cartografia preesistente e a scala superiore. Questa astrazione, fino a qualche anno fa, era svolta interamente a mano dal cartografo, richiedendo tempi lunghissimi. Con l'avvento al digitale, però, è stata intravista la possibilità di automatizzare questo processo, e di diminuirne conseguentemente tempi e costi. In questo ambito si va a collocare il progetto di ricerca CARGEN dell'Università di Padova, che mira a sviluppare un processo di generalizzazione automatico per la produzione di un database topografico in scala 1:25 000 e 1:50 000, partendo dalle carte tecniche regionali, fornite dalla Regione Veneto, in scala 1:5000.

Il presente lavoro di tesi si inserisce nel progetto CARGEN e si pone come obiettivo specifico lo studio dei problemi della toponomastica nella generalizzazione cartografica digitale e l'implementazione di possibili soluzioni per la vestizione toponomastica della carta (Label Placing), aspetto di cruciale importanza per la qualità del lavoro di generalizzazione e per la leggibilità di ogni tipo di mappa.

Il toponimo è il nome di un luogo legato al significato che questo ha avuto nella storia. Ecco allora che il toponimo si carica di un significato storico che diventa di grande importanza poiché è capace di dare delle informazioni preziose, quasi uniche, e non comunemente individuabili nei documenti.

L'intero lavoro è strutturato in otto capitoli. Nel secondo capitolo viene introdotto il tema della cartografia, definendo, in particolare, in cosa consiste il processo cartografico per la produzione delle mappe e come viene definita la generalizzazione cartografica, descrivendone alcuni modelli concettuali propo-

---

sti in letteratura. Il terzo capitolo è dedicato al progetto CARGEN, vengono delineati gli obiettivi del progetto e gli strumenti di lavoro maggiormente utilizzati. Il quarto capitolo propone una panoramica sul tema del Label Placement, analizzandone la storia, l'evoluzione, le difficoltà di risoluzione e le soluzioni proposte in campo internazionale sino ad oggi. Il quinto capitolo analizza nel dettaglio come venga trattata la toponomastica nella cartografia digitale, come venga generata, quali sono le regole da seguire e come esistano diverse filosofie ed approcci al suo trattamento. In particolare si presenterà la situazione in Italia.

Nel sesto capitolo sarà presentato l'approccio ad una soluzione concreta al problema del piazzamento di etichette su cartografia digitale realizzato in CARGEN ed implementato in Java. Si vedranno i concetti e la logica dietro alla soluzione proposta, gli algoritmi ed i risultati ottenuti. Il settimo capitolo integra il sesto mostrando i risultati ottenuti da test su dati cartografici reali, ottenuti dal portale Regionale del Veneto. Infine, l'ultimo capitolo analizza i possibili lavori e sviluppi futuri del tema sia in campo generale sia nello specifico del progetto CARGEN e della soluzione proposta in precedenza.

## 2. CARTOGRAFIA E GENERALIZZAZIONE CARTOGRAFICA

Per introdurre il tema della generalizzazione cartografica, in questo capitolo verranno, per prima cosa, fornite alcune definizioni relative ai concetti di cartografia, carta geografica e scala. Seguirà un breve resoconto sulla storia della cartografia. Successivamente, saranno illustrate in modo generale le varie classificazioni e tipologie delle carte geografiche e i sistemi informativi geografici (GIS), per concludere infine con la descrizione del processo cartografico, processo tramite il quale si realizza una mappa.

### 2.1 *La cartografia*

Per introdurre le basi della cartografia è necessario spiegare dei termini fondamentali alla comprensione dell'argomento nella sue generalità.

La *cartografia* è la scienza che si occupa di riprodurre l'aspetto della Terra o di una sua parte con la minima distorsione possibile, in modo che le informazioni contenute sulla carta siano il più possibile accurate.

---

*Enc. Italiana Grolier, 1987*

La *carta geografica* 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.

---

*Diz. Devoto, 1979*

La scala secondo cui la carta geografica rappresenta la realtà è il rapporto fra le distanze risultanti sulla scala stessa e le corrispondenti distanze reali sulla superficie della Terra. Così, Le carte su grande scala riproducono una piccola zona e possono contenere numerosissimi dettagli. Al contrario, le carte su piccola scala coprono un'area più vasta evidenziando un ristretto e selezionato numero di informazioni. Si può definire una classificazione delle

mappe basata sulla scala che metta in evidenza la relazione tra valore di scala adottato e informazioni trasmesse dalla mappa, e ricavare in base a questo criterio quattro gruppi principali di mappe:

- mappe a scala maggiore di 1:10 000, usate per le rappresentazioni catastali e per le carte tecniche dei centri abitati o delle aree rurali ed industriali, come la Carta Tecnica Regionale (CTR);
- mappe a scala compresa tra 1:10 000 e 1:100 000, carte topografiche rappresentanti con molta accuratezza zone relativamente limitate della superficie terrestre (rilievi, idrografia, orografia, edificato), come le carte dell'Istituto Geografico Militare;
- mappe a scala compresa tra 1:100 000 e 1:1 000 000, che sono carte corografiche rappresentanti, con buona ricchezza di particolari, aree abbastanza estese della superficie terrestre come regioni o intere nazioni;

La storia della cartografia, in Italia, inizia nel 1872, alla fondazione dell'Istituto Topografico Militare (diventato nel 1882 Istituto Geografico Militare), che nello stesso anno (1872) iniziò la rilevazione del territorio italiano per la realizzazione della nuova Carta Topografica d'Italia (1921) alla scala 1:100 000.

Ad oggi, si assiste ad una vera e propria rivoluzione cartografica, animata dall'impiego massiccio dell'informatica e dall'utilizzo di nuovi mezzi come i sistemi satellitari o le rilevazioni aeree. Nascono i Sistemi Informativi Geografici (GIS, Geographical Information System), potenti strumenti di analisi dei dati geografici, che permettono non solo la visualizzazione, ma anche l'elaborazione e la manipolazione degli elementi di una carta.

## 2.2 Il processo cartografico

Il processo cartografico, ovvero il processo di realizzazione di una carta, è il frutto di una sequenza ordinata di operazioni

- analisi e la definizione delle caratteristiche finali della mappa;
- raccolta dei dati;
- produzione della mappa e collaudo.

Tale processo resta valido sia nel caso in cui la mappa sia realizzata a mano dal cartografo sia nel caso in cui si faccia uso di tecniche e strumenti informatici.

La prima fase è quella che coinvolge la progettazione della mappa, ovvero la definizione delle caratteristiche tecniche e dei contenuti. Le caratteristiche

tecniche principali sono la scala, la superficie di riferimento, la proiezione e il tipo di rappresentazione (conforme, equivalente, equidistante). I contenuti della mappa, ovvero le informazioni che si vogliono collocare nella mappa, determinano il tipo di astrazione e rappresentazione che saranno utilizzate per trattare i dati in ingresso. In base alle decisioni prese a questo livello, la carta assumerà una natura tematica, cioè focalizzata solo su alcuni aspetti, piuttosto che una natura olistica, ovvero mirata a dare una visione d'insieme.

In questa prima fase si definisce, dunque, come debba avvenire l'astrazione della realtà, e in questo modo si produce anche, per le tradizionali mappe cartacee, la legenda della carta, e per le mappe digitali, la definizione di un GeoDB (Database digitale geografico).

Terminata la fase di analisi e definizione, si passa all'acquisizione dei dati. I dati possono essere raccolti in due modi, a seconda che la mappa sia il frutto di una rilevazione o di una derivazione. Nel primo caso, gli elementi del territorio vengono rilevati durante un sopralluogo con l'aiuto di sistemi di posizionamento satellitare (GPS) e fotografie aeree. Nel secondo caso, se la mappa viene derivata, vengono utilizzati dati provenienti da una cartografia preesistente, la quale dovrà avere un livello di dettaglio maggiore (e quindi un'alta scala) rispetto alla carta che si deve realizzare (questo è necessario in quanto solamente se si è in possesso di tutti i dettagli è possibile astrarre e rappresentare correttamente la realtà). In entrambi i casi, gli operatori dovranno assegnare ad ogni elemento individuato un codice diverso a seconda della natura dell'elemento (strada, ferrovia, edificio, fiume etc.), in accordo col modello già stabilito nella fase di analisi e definizione.

Una volta che il cartografo entra in possesso di tutte le informazioni (dati e specifiche) necessarie, avvia il processo di generalizzazione, fondamentale per la costruzione della mappa. Durante questa fase, il cartografo deve prendere una serie di decisioni atte a soddisfare le specifiche definite, ma anche a garantire i requisiti classici di ogni mappa, quali la leggibilità e l'usabilità. In pratica egli dovrà creare un'astrazione della realtà geografica che faciliti la comprensione e la comunicazione dell'informazione.

L'ultima fase del processo cartografico consiste nel collaudo della mappa, durante il quale si verifica la correttezza e la consistenza della carta. Tra le attività più importanti di collaudo vi è il controllo finale sul terreno mediante operazioni di misura e verifica della rappresentazione cartografica: vengono confrontati sul campo i dati riportati nella mappa con rilevazioni effettuate tramite strumenti ad alta precisione, come ad esempio il GPS differenziale.

### 2.3 I GIS (*Geographic Information Systems*)

Un GIS è una potente serie di strumenti per acquisire, memorizzare, estrarre alla bisogna, trasformare e visualizzare dati spaziali dal mondo reale.

---

*Burrough, 1986*

Con l'avvento dei GIS la cartografia subisce un cambiamento epocale. I GIS, infatti, permettono di manipolare oggetti spaziali e di estrarre, in vari modi, una grande quantità di informazioni dalle mappe: è così che la mappa si trasforma in una sorta di oggetto dinamico, mentre sino a poco prima era vista come un elemento statico. Il singolo dato geografico diventa un'entità dotata di attributi e inserita in una gerarchia relazionale; a questa entità corrisponde, oltre che una rappresentazione grafica, un contenuto logico e semantico.

In altre parole, un GIS permette di integrare operazioni che tipicamente offre un database con i vantaggi offerti dalla visualizzazione della mappa. Le entità geografiche, elementi base del GIS, vengono memorizzate all'interno di un database geografico (GeoDB), detto anche database spaziale, la cui funzione è quella di memorizzare, interrogare e manipolare informazioni geografiche e dati spaziali. Più precisamente, un sistema GIS è costituito da risorse e procedure. Le risorse coinvolte sono risorse umane, infrastrutturali e di dati. Le procedure, invece, si suddividono in acquisizione, archiviazione, elaborazione, presentazione e trasmissione.

I dati sono in assoluto l'elemento più prezioso del sistema, perché hanno un costo di produzione elevato. Essi, inoltre, per mantenere il loro valore, devono costantemente essere aggiornati. Ogni dato, che viene chiamato *feature*, è costituito da due componenti: la componente spaziale, che contiene informazioni relative alla posizione (geografica), alla geometria (forma e dimensione) e alla topologia (relazioni spaziali con altri dati) del dato; e la componente non spaziale, formata da dati descrittivi, o attributi, e metadati associati alla componente spaziale che velocizzano le operazioni di query spaziale.

L'acquisizione dei dati avviene normalmente scannerizzando le mappe cartacee tradizionali. Un altro metodo, invece, si basa sull'utilizzo di file contenenti 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. Ovviamente, al momento dell'acquisizione è necessario memorizzare i dati. Il modo in cui le informazioni spaziali vengono memorizzate dipende da quale modello dello spazio

viene adottato dal GIS. Ci sono due modelli utilizzabili: il modello raster e il modello vettoriale.

Nel primo caso, lo spazio viene visto come un mosaico, dove ogni tassello, o pixel, contiene uno o più valori (ad esempio, il colore o l'altitudine). Il pixel rappresenta il luogo dove una grandezza assume un certo valore. Un file raster risulta allora costituito da una matrice quadrata dove ogni cella è un pixel. Le dimensioni di tale file tendono ad essere molto elevate, in quanto ogni unità di spazio è sempre descritta da un pixel.

Il secondo modello, quello vettoriale, è formato da una serie di primitive geometriche che vengono istanziate all'interno di uno spazio vuoto. In questo secondo caso, a differenza di quanto avviene nel modello raster, lo spazio non viene riempito completamente, ma funge da contenitore di oggetti. Le primitive vettoriali che costituiscono il modello sono:

- il punto, definito da una coppia  $(x, y)$  o da una terna di coordinate  $(x, y, z)$ ;
- la linea, costituita da una sequenza ordinata di punti;
- il poligono, anch'esso costituito da una sequenza ordinata di punti, dove però l'ultimo punto della sequenza corrisponde al primo.

Il punto viene utilizzato quando l'elemento che si vuole rappresentare è semanticamente un punto, oppure quando, alla scala a cui si opera, esso è di fatto un punto. La medesima considerazione vale anche per la linea. Il poligono, invece, viene utilizzato per definire una particolare area della superficie geografica (lago, edificio, confine amministrativo).

Comparando il modello raster e quello vettoriale, emerge che quest'ultimo si presta molto di più alla manipolazione, all'elaborazione e all'interrogazione dei dati, che si possono integrare facilmente in un database relazionale, dove vengono trattati come valori di una colonna di una tabella, e processati per mezzo di molteplici operatori. Ciò nonostante, il GIS viene progettato in modo che possa supportare entrambi i modelli. A seconda di quale utilizzo viene fatto della carta, poi, sarà adottato un modello piuttosto che l'altro (ad esempio, le scansioni aeree o le immagini satellitari nascono e vengono utilizzate in formato raster).

In fase di stampa la mappa viene spesso convertita dal formato vettoriale a quello raster, per consentire stampe in grande quantità con procedure tipografiche. Una volta che i dati sono stati acquisiti e opportunamente memorizzati secondo uno dei due modelli appena illustrati, è possibile effettuare su di essi operazioni di elaborazione e trasformazione degli elementi geografici tramite degli strumenti di analisi forniti dal GIS. Alcune di queste operazioni sono:

*overlay topologico*: in cui si effettua una sovrapposizione tra gli elementi di due temi per creare un nuovo tematismo, ad esempio per sovrapporre il tema dei confini di un parco con i confini dei comuni in modo da determinare le superfici di competenza di ogni amministrazione o la percentuale di area comunale protetta;

*query spaziali*: ovvero interrogazioni dei dati a partire da criteri spaziali (vicinanza, inclusione, sovrapposizione);

*buffering*: permette di creare un poligono che circonda la geometria originaria;

*segmentazione*: che applicato su un elemento lineare, determina il punto alla distanza specificata dall'inizio dell'elemento;

*network analysis*: che applicata su una rete di elementi lineari, ad esempio la rete stradale, determina i percorsi minimi tra due punti;

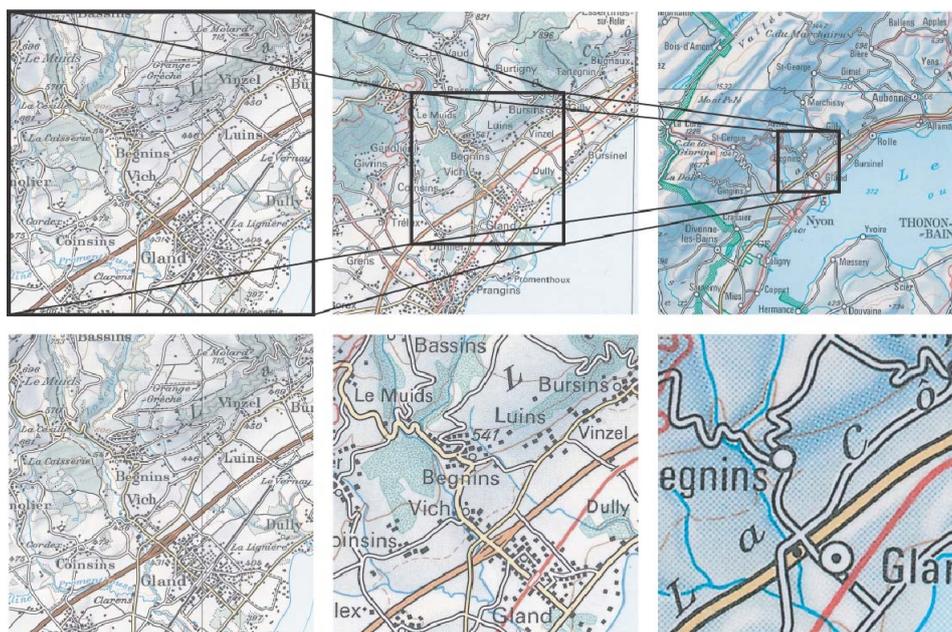
*spatial analysis*: che effettua un'analisi spaziale di varia tipologia utilizzando un modello raster, come ad esempio l'analisi di visibilità.

## 2.4 La generalizzazione cartografica

La generalizzazione è, senza alcun dubbio, uno dei passi chiave nella realizzazione di una mappa. 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 tutta la realtà. Al contrario, è necessario selezionare le informazioni rilevanti. Il risultato sarà dunque una astratta limitata parte della realtà, vedi Fig. 2.1.

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

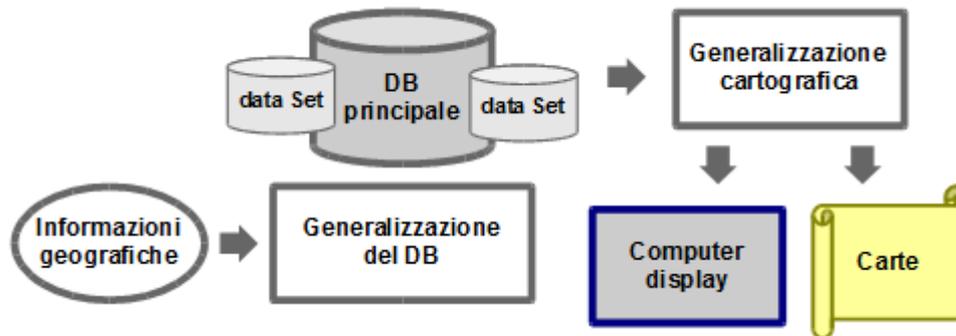


**Fig. 2.1:** esempio di generalizzazione di una carta dalla scala 1:100 000 alla scala 1:200 000 e 1:500 000. In alto, le dimensioni delle carte sono riportate coerentemente con il valore della scala. In basso, le tre carte sono riportate secondo la stessa distanza grafica. *Cecconi (2003)*

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) ne individuano sei:

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.

A seconda dei dati di partenza per la costruzione della mappa, si possono distinguere due processi di generalizzazione: se i dati disponibili sono il frutto di una rilevazione, cioè di un'acquisizione diretta, allora si parla di *compilazione cartografica*, se, invece, i dati sono quelli provenienti da una cartografia preesistente, si parla di *generalizzazione cartografica*.



**Fig. 2.2:** fasi della generalizzazione all'interno di un GIS: modellazione delle informazioni geografiche, che definisce il modello dati del GeoDBR, e generalizzazione delle feature. *Deruda (2005)*

#### 2.4.1 Il processo di generalizzazione

Il processo di generalizzazione può essere definito in questo modo:

Il processo di generalizzazione è la selezione e la rappresentazione semplificata dei dettagli che meglio si adatta alla scala e o allo scopo di una mappa.

*Burrough, 1986*

Due sono le componenti che caratterizzano questo processo (vedi Fig. 2.2: la generalizzazione del modello e la generalizzazione cartografica, che verranno esaminate in dettaglio nelle sezioni seguenti.

##### *Generalizzazione del modello*

Per mezzo della generalizzazione del modello, il risultato che si ottiene è la creazione di un modello astratto, rappresentativo della realtà. La scelta relativa a quali aspetti della realtà riportare e quali invece eliminare avviene in funzione degli scopi per cui viene realizzata la mappa. Ad esempio, in una carta politica verrà dato risalto ai confini amministrativi mentre elementi fisici come i rilievi o le depressioni non verranno riportati.

Questo processo porta alla definizione di una tassonomia degli elementi della realtà, cioè ad una loro suddivisione in classi. Anche i parametri che portano a distribuire gli oggetti in più classi variano a seconda delle funzioni della mappa che si vuole realizzare: così, una carta civile può classificare le strade in urbane ed extraurbane, mentre una carta militare può enfatizzare

maggiormente una distinzione delle vie di comunicazione basata sulla loro larghezza e tipo di fondo.

Nella cartografia digitale, la generalizzazione del modello corrisponde alla definizione di un GeoDatabase. In fase di definizione del database, è fondamentale ricordare che una modellazione e strutturazione efficiente della base di dati rappresenta un presupposto necessario per la buona riuscita della fase di generalizzazione cartografica, nonché per un proficuo utilizzo della cartografia digitale in ambiente GIS.

### *Generalizzazione cartografica*

La seconda componente della generalizzazione è quella cartografica, che agisce sulle informazioni geografiche delle *feature*.

Una volta definito il modello astratto della realtà, si passa alla ridefinizione delle geometrie, allo scopo di fornire una coerente rappresentazione grafica al modello dei dati. Questa fase è di fondamentale importanza per garantire una mappa accurata e comprensibile. Il cartografo, per ridefinire le geometrie, si deve affidare alla sua interpretazione personale: non esiste infatti nessuno schema rigido che si possa applicare sistematicamente ad ogni singolo caso. A titolo d'esempio, un agglomerato di case potrebbe venire rappresentato come un unico blocco da un primo operatore e da due o più blocchi da un secondo operatore.

Tra i fattori che possono influenzare le scelte compiute dal cartografo troviamo la scala di rappresentazione, e, successivamente il contesto. In relazione al valore di scala, un centro urbano potrebbe venire rappresentato con un punto affiancato dal nome della città oppure da un suo sottoinsieme di strade ed edifici. Il contesto invece determina il livello di importanza attribuito ad un oggetto, che può determinare la sua eliminazione o la precisione con cui lo si rappresenta. Ad esempio, un'etichetta posizionata in una zona densamente edificata potrebbe essere omessa, per garantire la massima leggibilità della mappa.

#### *2.4.2 Generalizzazione automatica*

Le tecnologie informatiche hanno permesso di automatizzare la procedura di generalizzazione, riducendo 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 di un foglio 1x121cm della serie Topo100<sup>1</sup> con

---

<sup>1</sup> Carte topografiche in scala 1:100 000.

i metodi tradizionali, l'IGN<sup>2</sup> 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 le mappe relative all'intero territorio francese. Diminuendo il tempo di produzione, invece, si possono aggiornare le mappe con maggior frequenza e minor sforzo 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 perché il campo della generalizzazione automatica desta sempre più interesse presso gli enti cartografici internazionali, stimolando la produzione scientifica in questo settore.

La realizzazione di un processo di generalizzazione automatico si deve comunque scontrare con alcuni ostacoli (Mackaness, 2007):

- *Complessità del processo di progettazione*, che non può prescindere dalla valutazione di molteplici fattori e vincoli: 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.

### 2.4.3 Modelli di generalizzazione

La generalizzazione, dagli anni Settanta 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 Settanta, Robinson (1978) sviluppò uno dei primi modelli concettuali, allo scopo di

---

<sup>2</sup> Institut Géographique National, <http://www.ign.fr>.

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, e 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 e posizionamento dei nomi (*name placing*). 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:

Why : perché generalizzare? Vengono individuati degli obiettivi filosofici (elementi teorici, elementi computazionali ed elementi specifici);

When : quando generalizzare? Si specifica una valutazione cartometrica delle condizioni che motivano la generalizzazione (condizioni geometriche, misure spaziali e olistiche, controllo delle trasformazioni)

How : come generalizzare? 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 AGENT<sup>3</sup>. 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.

---

<sup>3</sup> Automated GEneralization New Tecnology, (AGENT, 2003), <http://agent.ign.fr/>

### 3. 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 (IGM). L'obiettivo di questo progetto è arrivare ad un processo di generalizzazione cartografica automatizzato applicabile alla produzione di mappe derivate. I dati di ingresso del processo sono quelli della cartografia tecnica regionale (CTR), alla scala 1:5000, mentre la cartografia che si intende produrre è quella coerente con i modelli DB25 e DB50 dell'IGM. In questo capitolo si descriverà la situazione cartografica attuale italiana, e si discuteranno gli obiettivi del progetto CARGEN. Sarà poi presentato l'ambiente di lavoro generale, dati e software utilizzati nel progetto.

#### 3.1 *La situazione della cartografia in Italia e CARGEN*

In Italia, l'ente che si occupa di mantenere una cartografia aggiornata a livello nazionale è l'IGM. La prima produzione cartografica dell'istituto è stata la Nuova Carta Topografica d'Italia, realizzata alla scala 1:100 000. L'IGM ha prodotto diverse carte che coprono il territorio a livello nazionale e solo una parte di queste vengono mantenute aggiornate.

Per la scala 1:25.000 esistono le serie 25V, 25 e 25DB, per la scala 1:50 000 le serie 50 e 50/L e per la scala 1:100 000 le serie 100V e 100L. Tra le varie serie sussiste un preciso rapporto matematico: ogni foglio di una mappa in scala 1:100 000 è divisa in quattro settori, rappresentati in altrettante mappe della serie 1:50 000. Queste mappe vengono dette *quadranti*, e ogni quadrante è a sua volta diviso in quattro parti, le *tavolette*, che sono in scala 1:25 000. Delle serie cartografiche alla scala 1:25 000, solo la più antica, la 25V (vecchio taglio), è stata completata. I dati di questa serie sono aggiornati mediamente al 1960, ad eccezione di alcuni elementi, che sono stati prodotti durante una campagna di aggiornamento parziale del 1984.

La serie 25 copre, dunque, solo il 36% del territorio nazionale, la sua produzione è stata interrotta per fare spazio alla nuova serie 25DB. Questa serie introduce la *cartografia digitale*, e, su un totale di 2298 sezioni, ne sono

state realizzate 68, ottenute tramite stereorestituzione<sup>1</sup> numerica o come derivazione dalla cartografia tecnica regionale numerica.

Oltre all'IGM, che lavora su cartografie con scala 1:25 000 e inferiori, vi sono altri enti che si occupano di cartografia e ai quali vengono affidate le realizzazioni di carte a media e grande scala (1:10 000 e superiore). Inizialmente era il Catasto a dover adempiere a questo compito, successivamente tale funzione è stata trasferita alle Regioni che hanno potuto così gestire in modo autonomo la creazione delle carte regionali, definite tecniche in quanto create specificamente per i tecnici delle amministrazioni. Le carte prodotte dalle Regioni sono carte ricche di particolari e vengono aggiornate frequentemente, operazione facilitata dalla ristretta porzione del territorio nazionale che viene rappresentata. La *Carta Tecnica Regionale (CTR)* costituisce la base di riferimento per la redazione degli strumenti urbanistici comunali, per i Piani di Coordinamento Provinciali, per i Piani d'Area e per i vari piani di settore della pianificazione e della programmazione regionale. In Veneto vengono prodotte due serie di CTR: una in scala 1:10 000, composta da sezioni, e una in scala 1:5000, composta da elementi. Il taglio e l'inquadramento le rendono sovrapponibili alle carte IGM di nuova produzione. In particolare, ogni foglio IGM in scala 1:50 000 è diviso in 4 quadranti della serie 25, ognuno dei quali ripartito in 4 sezioni in scala 1:10 000 della CTR, divisi a loro volta in 4 elementi in scala 1:5000.

Recentemente anche gli uffici cartografici regionali hanno intrapreso la strada verso la digitalizzazione cartografica, iniziando la produzione di *Carte Tecniche Regionali Numeriche (CTRn)*. Questo fatto, in concomitanza con il passaggio, da parte dell'IGM, dalla serie 25 alla 25DB, ha aperto un nuovo scenario nella cartografia italiana: la possibilità di usare la derivazione come mezzo per produrre la nuova serie 25DB. Questa situazione permetterebbe quindi di limitare le rilevazioni dei dati sul territorio, e accelerare i tempi di produzione delle sezioni della serie 25DB. La realizzazione di una procedura informatica per la generalizzazione cartografica accelererebbe ulteriormente i tempi, dando la possibilità di automatizzare il processo di derivazione. Il progetto CARGEN nasce proprio per questo scopo, ovvero sviluppare un processo di generalizzazione delle carte tecniche regionali per produrre una base di dati coerente con il modello 25DB dell'IGM.

---

<sup>1</sup> La *stereorestituzione* è un'operazione che, a partire da una coppia di fotogrammi che riprendono con angolazioni diverse la medesima zona, ne ricostruisce lo spazio virtuale.

### 3.2 Ambiente di lavoro

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 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. 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 Java sviluppata all'interno del Progetto. Inoltre, una libreria potentissima sempre Java, la *JTS Topology Suite*<sup>2</sup>, fornisce una vasta serie di operatori spaziali, evitando così di ricorrere al DBMS di Oracle per effettuare le interrogazioni spaziali. Nei successivi sottoparagrafi saranno brevemente descritti i software più importanti utilizzati all'interno del progetto. Per una descrizione approfondita e completa degli inizi e degli sviluppi del progetto CARGEN si rimanda alla lettura della tesi di laurea [vedi 12] del dottorando Sandro Savino, presente nel progetto dal 2006, anno della nascita.

#### 3.2.1 *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'involucro convesso;

---

<sup>2</sup> Realizzata da Vividsolutions, <http://www.vividsolutions.com/jts/JTSHome.htm>

- alcune funzioni per la semplificazione delle geometrie, come l'algoritmo di Douglas-Peucker;
- la funzione per la costruzione del Minimum Bounding Box.

Oltre a queste funzioni, la JTS fornisce l'implementazione di indici spaziali, come il *quadtree*<sup>3</sup>, che offrono un modo veloce per la risoluzione di query spaziali.

### 3.2.2 OpenJump

OpenJump<sup>4</sup> 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.

Un plugin da integrare offre al programmatore il grosso vantaggio di poter visionare tramite l'interfaccia grafica gli effetti della propria applicazione e i risultati dei propri algoritmi. Il plugin diventa così uno strumento essenziale nello sviluppo di nuovi algoritmi che manipolano geometrie: avere un riscontro grafico istantaneo è un grande aiuto per semplificare e velocizzare la fase di testing e debugging. 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.

---

<sup>3</sup> Struttura dati ad albero nella quale tutti i nodi interni hanno esattamente quattro nodi figli. Esso viene utilizzato per partizionare lo spazio bidimensionale, suddividendolo ricorsivamente in quattro quadranti.

<sup>4</sup> <http://www.openjump.org/>

## 4. IL PROBLEMA DEL LABEL PLACEMENT

In questo capitolo si andrà ad approfondire il tema principale da cui deriva il presente lavoro di tesi, ovvero il piazzamento delle etichette in modo automatico su mappe, il *Automatic Label Placement*, problema molto vasto e ampiamente dibattuto, ancora oggi oggetto di studi e applicazioni in molti campi (vedi Fig. 4.1).

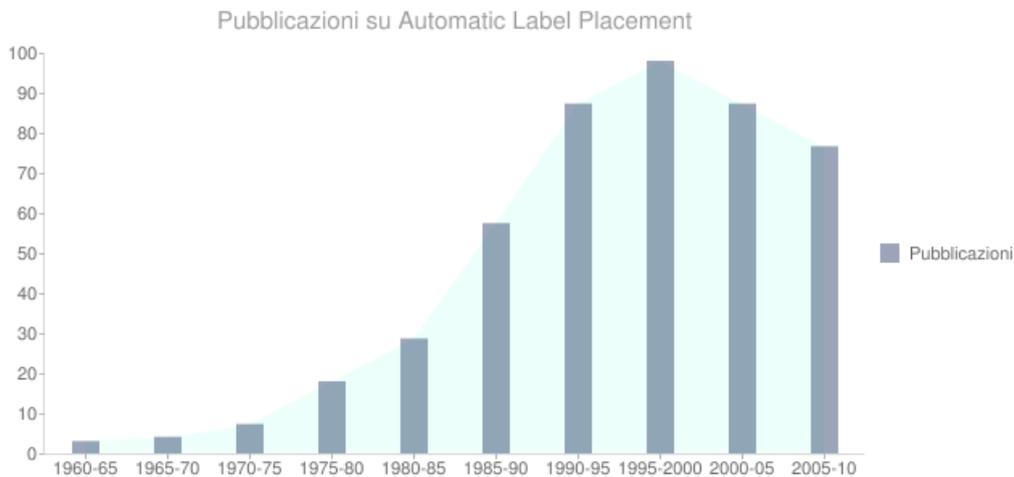
Saranno elencate le difficoltà della creazione di applicazioni pratiche che tentino di risolvere il problema e si darà una panoramica sulla vasta letteratura presente a riguardo.

### *4.1 Introduzione all'Automatic Label Placement e problematiche annesse*

Il problema del piazzamento automatico delle etichette si riferisce a metodi automatici generati da calcolatori elettronici per il piazzamento di nomi ed etichette su carte, mappe e grafici. In una mappa si possono vedere tre tipi di oggetti: linee, aree e punti. E' di fondamentale importanza l'etichetta da associare ad ognuno di questi elementi in modo che si capisca immediatamente a quale oggetto l'etichetta faccia riferimento. Infatti un buon piazzamento delle etichette migliora la leggibilità di una mappa tanto quanto il giusto posizionamento degli oggetti visivi all'interno dell'area definita.

Il posizionamento automatico del testo è uno dei problemi più difficili, complessi e dispendiosi in termini di tempo nella creazione di mappe e nei GIS. Il problema, infatti, è noto per essere NP-Difficile[10]. Esistono vari parametri da tenere in considerazione nell'ottimizzazione del piazzamento di etichette in una mappa, e i principali punti da trattare sono due (Marks e Shieber [10, p. 2]):

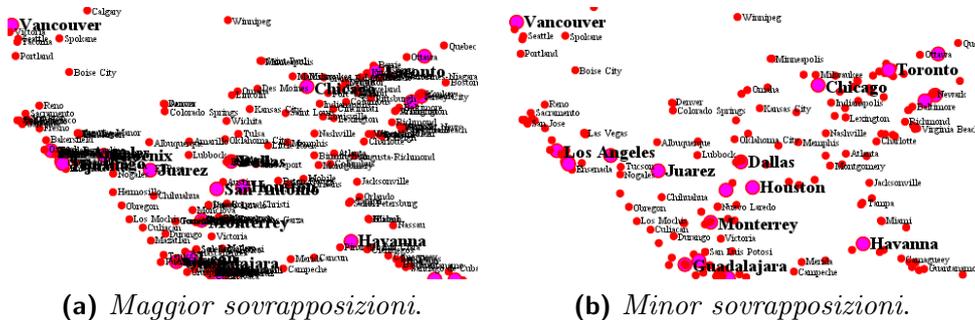
1. In quale grado le etichette si sovrappongono e oscurano elementi cartografici della mappa, incluse altre etichette.
2. In quale grado le etichette sono leggibili e facilmente associabili all'elemento cartografico che rappresentano senza ambiguità.



**Fig. 4.1:** Il grafico illustra l'andamento del numero delle pubblicazioni internazionali ufficiali sul tema dell'Automatic Label Placement. Come si può notare è un tema centrale anche al giorno d'oggi. Fonte: <http://i11www.iti.uni-karlsruhe.de/~awolff/map-labeling/>

Come si può notare in Fig. 4.2 il posizionamento dei nomi è cruciale nella leggibilità di una mappa. Ma eliminare le sovrapposizioni spesso non è sinonimo di qualità. Infatti l'azione di cancellazione di un'etichetta va ben ponderata, potrebbe compromettere il riferimento ad elementi importanti all'interno della mappa o grafico da stampare!

Ad esempio, se il nome di un piccolo fiume si sovrappone a quello di una importante città, sarebbe più opportuno omettere il primo per mantenere un grado di informazione e qualità il più alto e adeguato possibile. Questo tipo di scelte, nella redazione di mappe geografiche e, in particolare, nel disegno di cartografie importanti, è sempre stato fatto da un cartografo professionista, figura umana fondamentale per l'interpretazione di caratteristiche astratte come la qualità estetica e la qualità d'informazione di una mappa. Ovviamente questo lavoro è molto dispendioso in termini di tempo, perciò, oggi, grazie alla tecnologia disponibile e agli studi nel campo della cartografia digitalizzata, si tenta di cedere il compito del piazzamento delle etichette ad applicazioni informatiche basate su algoritmi di scelta complessi. I risultati sono promettenti ma ancora di scarsa praticità. Questo è dovuto al fatto che ancora non esistono regole internazionali ben definite che fissino degli standard sulla trattazione della toponomastica nelle mappe. Ci sono varie possibili soluzioni per il piazzamento automatico delle etichette, basate su

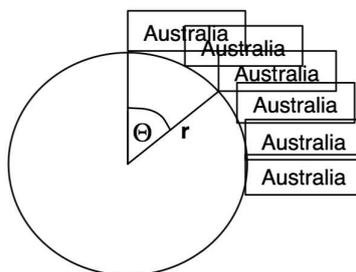


**Fig. 4.2:** Si può notare il diverso grado di leggibilità delle due mappe. Purtroppo l'eliminazione delle etichette che si sovrappongono deve essere ponderato e deve considerare l'importanza degli elementi associati alle etichette.

algoritmi di vario tipo:

- *Algoritmi basati su regole:* sono algoritmi complessi che, tramite un insieme di regole comportamentali, tentano di simulare l'approccio umano di un cartografo. Questo permetterebbe ad esempio di stampare sul mare nomi di città costiere, ripetere il nome di un fiume più volte lungo il suo corso, ecc.
- *Algoritmi greedy:* piazzano etichette consecutive in modo che non si sovrappongano. I risultati sono poco soddisfacenti anche per piccole mappe, ma sono estremamente veloci.
- *Algoritmi di ottimizzazione:* spostano un'etichetta alla volta finché questa migliora la sua posizione. Arrivano ad una soluzione di ottimo locale. Sono soddisfacenti soprattutto in mappe a bassa densità.
- *Altri:* esistono molti altri tipi di algoritmi, come *Simulated Annealing* (evitano il problema dell'ottimo locale), *Algoritmi Evolutivi*, più complessi e che garantiscono risultati migliori.

Una delle possibilità per la riduzione della difficoltà del piazzamento dei nomi è quella di dividere la mappa in zone diverse ognuna indipendente dalle altre. Facendo questo in modo ricorsivo si potrebbe arrivare ad avere insiemi sempre più piccoli fino ad arrivare a risolvere sovrapposizioni fra sole due etichette. In questo caso la soluzione sarebbe data dalla risoluzione di un'istanza di *2-satisfiability*. Molti algoritmi sono fondati su questo principio. Per un'analisi approfondita dei problemi e delle possibili soluzioni teoriche legate al *Map Labeling* si veda il lavoro di Itarriaga [9].



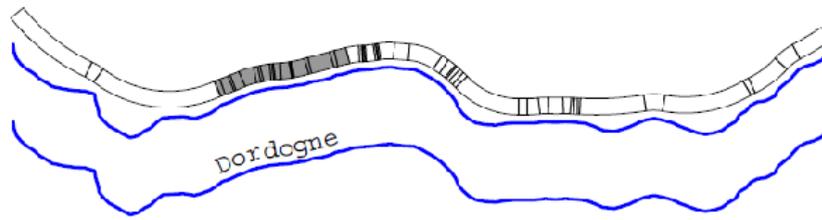
**Fig. 4.3:** Nell'immagine si osserva come viene posizionata l'etichetta di un punto nel lavoro di Barone et. al.

## 4.2 Panoramica sulla letteratura

Nonostante l'importanza internazionale del tema, non esistono soluzioni pratiche che garantiscano risultati soddisfacenti, ovvero che generino automaticamente etichette applicate con qualità sulle mappe. Questo è dovuto principalmente alla categoria del problema da affrontare, che, ricordiamo, è NP-Difficile.

Le prime soluzioni concrete basate su algoritmi arrivano negli anni Ottanta, grazie anche al lavoro di Ahn e Freeman [vedi 1], i quali proponevano un algoritmo basato su regole intelligenti per etichette da posizionare su punti, linee e aree. Questi algoritmi di ottimizzazione garantivano soluzioni di qualità però richiedevano tempi di esecuzione troppo alti. Col passare degli anni si sono tentati altri approcci per migliorare il tempo d'esecuzione degli algoritmi, come quello di Edmondson e Christensen [vedi 5], i quali hanno proposto un algoritmo ibrido basato su euristiche cartografiche e principi di ottimizzazione per raggiungere sia qualità che efficienza. I risultati migliori e gli algoritmi più efficienti sono stati ottenuti nel caso di *point feature label placement problem (PFLP)*, ovvero piazzamento di etichette associate a punti. Uno dei migliori esempi è dato dal lavoro di Lecchi Cravo et al.[4]. L'algoritmo utilizza una procedura greedy intelligente denominata GRASP (Greedy Randomized Adaptive Search Procedure) che si dimostra più efficace di tutti gli algoritmi precedentemente mostrati in letteratura basati su ottimizzazione. Sempre per il problema del point labeling, Bradstreet, Barone e White hanno introdotto un algoritmo evolutivo [3], tipico delle applicazioni di ricerca nel campo della genetica, che amplia il numero di posizioni possibili per un'etichetta (vedi Fig. 4.3), e sfrutta un sistema di base multi-oggetto per i processi di decisione. Di importanza rilevante anche l'algoritmo basato su tabu-search realizzato utilizzando il framework POPMUSIC[2], sempre per PFLP.

Per quanto riguarda la branca del *Line Labeling*, è notevole il lavo-



**Fig. 4.4:** Si nota come nel lavoro di Wolff l'etichettatura su polilinee venga gestita in modo molto preciso ed efficace.

ro di Alexander Wolf et. al [14] dove si illustra un efficace algoritmo di posizionamento di etichette su polilinee, illustrato brevemente in Fig. 4.4.

Molti altri algoritmi sono stati presentati ma la maggior parte sono di scarso utilizzo pratico. Rimangono valide proposte teoriche ma ancora irrealizzabili per limiti tecnici delle macchine odierne. Esiste anche una libreria scritta in C++ e Java di algoritmi di ottimizzazione e generazione di etichette automatizzate che sta crescendo con ottime aspettative. Si chiama PAL [vedi 6].

Recentemente è stata presentata una soluzione semi-automatizzata, chiamata *User Hints*[11]. Si tratta di un framework dove il cartografo viene aiutato da alcuni algoritmi di posizionamento nell'esecuzione del compito del piazzamento dei nomi. In sostanza non esiste un solo algoritmo complesso che automatizzi il piazzamento e le scelte, ma una serie di algoritmi più semplici, uniti ad un'applicazione grafica snella, che aiutano e accompagnano un professionista alle scelte migliori. Questo approccio garantisce un minor tempo di produzione rispetto alla generazione completamente manuale e riduce gli ovvi errori dati da una generazione completamente automatizzata.

Per un riassunto generale sugli sviluppi del tema e su tutte le soluzioni finora adottate è utile consultare il lavoro di Wolff [13], elaborato del 1999.

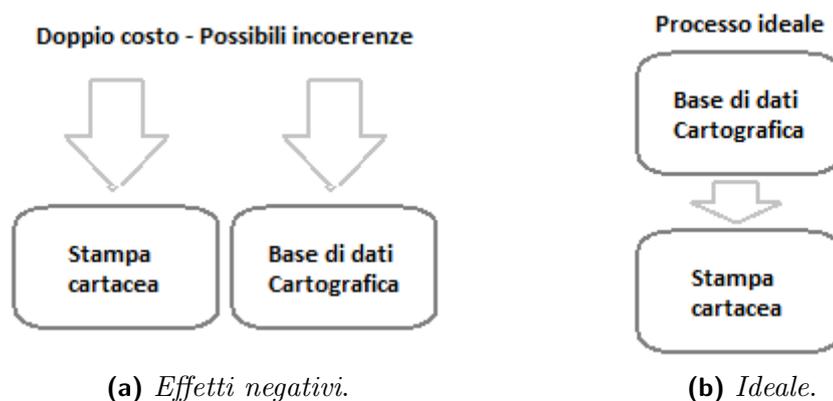
## 5. IL TRATTAMENTO DELLA TOPONOMASTICA NELLA CARTOGRAFIA DIGITALE

In questo capitolo sarà presentata la situazione internazionale per quanto riguarda il trattamento della toponomastica nella cartografia digitale. Non esiste uno standard internazionale nè tantomeno nazionale che definisca dei criteri per il trattamento della stessa. Si andrà quindi a presentare le difficoltà che derivano dalla mancanza di uno standard, si vedrà come attualmente venga trattato il problema dai vari software e le varie ideologie sul trattamento dei toponimi, findendo per illustrare come la toponomastica sia gestita in Italia, in particolare come l'IGM si comporta a riguardo.

### 5.1 *Introduzione al problema*

I toponimi condividono, con i confini, la caratteristica di essere oggetti della carta in qualche modo giustapposti alla realtà rappresentata. D'altra parte toponomastica e limiti amministrativi sono certamente indispensabili alla lettura della carta stessa. Questa sezione intende presentare le problematiche relative alla gestione dei toponimi nei GIS e nella cartografia digitale in genere. Le idee sul trattamento degli oggetti geometrici (puntuali, lineari, areali) si sono quasi del tutto chiarite nel corso degli ultimi anni, mentre la gestione digitale dei toponimi, invece, è ancora ben lungi dall'essere definita in modo uniforme. Allo stato attuale delle cose, i vari prodotti commerciali, ma anche le specifiche internazionali, adottano strategie non uniformi. E' importante definire cosa sia, per un toponimo, l'informazione geometrica, l'informazione attributo, e l'informazione di presentazione grafica.

Il fatto che i diversi GIS o Database cartografici utilizzino rappresentazioni e regole diverse per la rappresentazione della toponomastica causa diversi problemi al momento della generazione della mappa. In sostanza, una volta scelto un particolare prodotto software, è questo che detta la filosofia di gestione dei toponimi al cartografo, a partire dalle sue limitazioni, e non viceversa come dovrebbe essere. Ci sono gravi problemi di importazione/esportazione fra vari formati di memorizzazione, per cui la scelta del formato diventa significativa, mentre dovrebbe essere trasparente all'utilizzo. Infine spesso capita che non



**Fig. 5.1:** La mancanza di regole e dati ben strutturati per la toponomastica nel database e la discrepanza con i software utilizzati porta ad un doppio lavoro per la generazione della mappa.

sia possibile utilizzare lo stesso strumento software per la stampa e per la produzione del database cartografico, per cui è necessario predisporre una doppia linea produttiva, che moltiplica i costi di produzione (vedi Fig. 5.1).

## 5.2 Come sono trattati i toponimi nei database cartografici

Gli oggetti della carta devono essere suddivisi e raggruppati. E' chiaro a tutti ormai come gli oggetti puntuali, lineari ed areali debbano essere trattati nella cartografia digitale: questi elementi geometrici sono raggruppati in feature class (classi di caratteristiche), che formano gli strati tematici della carta. Ad ogni oggetto è associata una ben nota forma geometrica (punto, linea od area) ed i relativi attributi della base di dati. Per i toponimi invece non esiste una filosofia di gestione condivisa da tutti. Sembra quasi che la toponomastica sia un aspetto minore della cartografia, un dettaglio secondario rispetto agli oggetti geometrici principali. Non è chiaro nemmeno se i toponimi debbano essere oggetti geometrici (feature class) a sé stanti, oppure siano etichette generate dall'attributo nome di altri oggetti: uno dei formati cartografici più diffusi (il formato Shape), ad esempio, non permette di definire i toponimi come feature class vere e proprie. L'idea che il toponimo sia l'attributo nome di un altro oggetto geometrico della carta (punto, linea od area) può portare ad alcuni inconvenienti, infatti molto spesso non esiste una relazione biunivoca fra toponimo e oggetto geometrico:

- *toponimi multipli per lo stesso oggetto*: un singolo fiume, descritto da un singolo oggetto geometrico lineare, avrà, lungo il suo corso, il proprio

toponimo ripetuto molte volte.

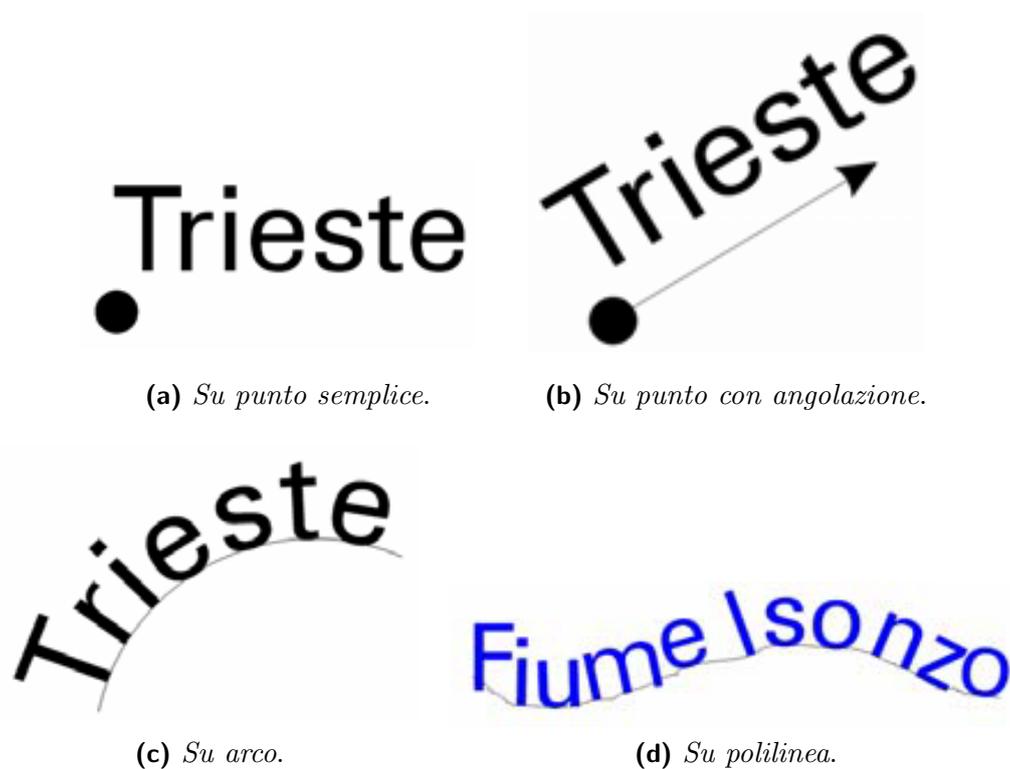
- *singolo toponimo per collezione di oggetti*: il toponimo di una città di una carta topografica sarà associato all'insieme degli oggetti geometrici che formano la città (case, ospedali, monumenti, parchi, etc.) e non corrisponde all'attributo nome di un particolare oggetto.
- *toponimo non associato a nessun oggetto*: il toponimo di una regione geografica (es. Mugello) non corrisponde a nessun oggetto geometrico.

Altri sistemi GIS prevedono invece che i toponimi siano oggetti geometrici propri, vale a dire feature class geometriche al pari di punti, linee ed aree. Questa filosofia (più europeo-mediterranea che anglosassone), prevede che i toponimi catturino l'informazione geometrica della loro posizione ed enfatizzano quindi l'aspetto artistico-interpretativo del cartografo che posiziona manualmente il toponimo (si ricorda che anche nei GIS come nella cartografia tradizionale, rimane sempre l'aspetto interpretativo-artistico del cartografo: si pensi ad esempio ad una serie di tornanti di montagna in una carta a grande denominatore di scala, in cui il cartografo deve rendere l'idea della serie di tornanti reali utilizzando solo poche curve).

### 5.3 La geometria dei toponimi

Per gli oggetti puntuali, lineari ed areali di una mappa, esiste uno standard di gestione comune a tutti i sistemi GIS (a parte qualche piccolo dettaglio). Per quanto riguarda i toponimi, invece, non c'è ancora uno standard per la definizione geometrica. Non si è raggiunto un accordo su come rappresentare geometricamente il toponimo, una volta deciso che questo è un oggetto appartenente ad una feature class. Le principali interpretazioni sono (vedi Fig. 5.2):

- *punto semplice (o etichetta automatica)*: il toponimo è rappresentato da una coordinata geografica, oppure è generato automaticamente come etichetta. In questo caso i toponimi sono visualizzati orizzontali.
- *punto più orientamento*: oltre all'informazione di posizione è presente l'angolo di orientamento. Questa interpretazione è quella più diffusa nei sistemi di stampa e nei software di tipo CAD.
- *arco di cerchio*: alcuni sistemi GIS avanzati permettono di spalmare il toponimo su di un arco di cerchio. Questa interpretazione è poco diffusa, quindi problematica da utilizzare. Alcuni software, che implementano



**Fig. 5.2:** Esempi di come i toponimi digitali possano essere stampati. Rispettivamente, a partire da un punto con scritta orizzontale, a partire da un punto con una determinata angolazione, su tutta la lunghezza di un arco di circonferenza, sulla lunghezza di una polilinea.

solo i toponimi come punti orientati, permettono di simulare gli archi di cerchio spezzando il toponimo in lettere singole.

- *polilinea*: il toponimo è spalmato su di una polilinea di forma qualsiasi. Questa interpretazione è la più espressiva dato che permette di simulare tutte le altre, ma è comunque molto rara (anche se supportata dal VPF).

#### 5.4 Forma e contenuto

I dati cartografici digitali, sia nella loro forma di sistemi informativi che, più semplicemente, in forma di cartografia digitale di supporto alla stampa

cartacea, sono caratterizzati da tre elementi: informazione geometrica, attributi associati, stile di visualizzazione o stampa. Per una strada ad esempio, l'insieme delle coordinate cartografiche che ne definisce la forma fa parte del primo punto (geometria), gli attributi come la classifica della strada e il tipo di fondo fanno parte del secondo (attributi associati), mentre il colore e lo spessore di stampa fanno parte del terzo (stile di visualizzazione). Al solito, per quanto riguarda la toponomastica, le idee sulla suddivisione fra informazione e stile di visualizzazione non sono chiare. Alcuni sistemi ad esempio interpretano la giustificazione (posizione relativa del toponimo) come informazione geometrica e quindi contenuto informativo, altri invece vedono la giustificazione come stile di visualizzazione. Inoltre, mentre per quasi tutti i sistemi la dimensione del font è sicuramente stile di visualizzazione, nei rari sistemi che supportano i toponimi come polilinee (esempio il VPF) la dimensione del font dipende dalla lunghezza della polilinea e quindi diventa contenuto informativo.

### 5.5 La trattazione con i software commerciali

Viene riportata di seguito una breve panoramica di come i software commerciali più diffusi trattano la toponomastica. Si noti in particolare la notevole diversità di gestione fra un software ad un altro.

- *Arcview*: nessuna gestione dei testi come feature. Possibilità di creare etichette a partire da attributi testo di altre feature. Dimensione, giustificazione e font fanno parte dello stile.
- *ArcGIS*: le geometrie standard non prevedono i toponimi. Esiste un livello speciale annotation con cui è possibile inserire testi.
- *Autocad*: inserimento di oggetti testo come punti orientati (non testi arcuati), la posizione può essere bi o tridimensionale. La giustificazione fa parte della geometria, il font e la dimensione del font fanno parte dello stile. Possibilità di simulare toponimi arcuati con appositi tools che frammentano il toponimo in oggetti composti dalle singole lettere orientate. I toponimi originali possono essere in qualche modo identificati usando particolari attributi (es. il gruppo grafico), ma risultano comunque di difficile gestione. Ad esempio l'ordine delle lettere viene perduto rendendo impossibile un'eventuale correzione a posteriori.
- *Microstation*: inserimento di oggetti testo come punti orientati (non testi arcuati), la posizione può essere bi o tridimensionale. La giustificazione

fa parte della geometria, il font e la dimensione del font fanno parte dello stile. Stessa possibilità di stampa di toponimi arcuati di Autocad

- *Geomedia*: possibilità di inserire oggetti testo come punti tridimensionali orientati, la giustificazione fa parte della geometria, la dimensione e il font fanno parte dello stile. In Geomedia 5.x sono stati introdotti anche testi come archi di circonferenza.
- *Dynamo*: inserimento di oggetti testo come punti (non orientati), oppure polilinee (possibilità quindi di disporre i testi su di una qualsiasi spezzata). Impossibilità di inserire testi come punti orientati semplici. La dimensione del font fa parte della geometria (su questo punto Dynamo è probabilmente unico), il font fa parte dello stile.
- *VPF*: possibilità di inserire oggetti testo come punti non orientati o come polilinee bi e tridimensionali. Non essendoci un sistema standard di visualizzazione, in questo contesto lo stile (font e dimensione) non ha significato.
- *Basi di dati*: nelle basi di dati con supporto geografico (Oracle 9-10, Postgresql-PostGIS, MySql 8.0 Spatial) nessun database geografico prevede un supporto specifico per i toponimi. Comunque si possono ricavare etichette a partire da attributi di altri feature (puntuali, lineari o areali). Non essendoci un sistema standard di visualizzazione, in questo contesto lo stile non ha significato.

### 5.6 I toponimi nel database secondo IGM

L'IGM ha un'idea diversa rispetto alla maggior parte dei GIS e delle comunità internazionali. Secondo l'istituto italiano, infatti, i toponimi devono avere queste caratteristiche:

- I Toponimi sono oggetti!
- La scelta attributo-etichetta è più legata al mondo anglosassone.
- L'aspetto interpretativo del toponimo (come tutto il resto della carta), non è derivabile con algoritmi automatici.
- Possono esistere toponimi non associati ad oggetti difficilmente identificabili (Regioni Geografiche).
- I toponimi sono punti orientati o polilinee.

- Il testo del toponimo non è un attributo ma una geometria.

L'IGM non cerca la completa automatizzazione del label placing, ma bensì vuole sfruttare possibili processi per aiutare il cartografo nel posizionamento manuale dei toponimi sulla carta, vista la loro rilevante importanza e, soprattutto, vista l'ovvia difficoltà interpretativa del toponimo data dai limiti di un software elettronico.

L'IGM produce due basi di dati legate alla toponomastica:

- *Il DB25 (e suoi derivati)*: la base di dati cartografici che contiene anche la toponomastica.
- *Il Database della Toponomastica*: è una base di dati classica contenente l'elenco di tutti i toponimi che compaiono nella cartografia IGM, con indicata la posizione indicativa e la tipologia del toponimo.

## 6. APPROCCIO DI POSIZIONAMENTO ETICHETTE SU CARTOGRAFIA DIGITALE IN CARGEN

In questo capitolo si andrà a presentare il lavoro svolto all'interno del progetto CARGEN per la realizzazione di una visualizzazione intelligente di etichette relative alla toponomastica su tutte le scale. Verranno inizialmente descritti gli strumenti di lavoro e le intenzioni del progetto di tesi. Successivamente verranno descritte le idee di base del lavoro e dell'algoritmo che genera la visualizzazione delle etichette cercando di evitare eventuali collisioni e che tiene conto del grado di densità di vertici (elementi) presente al di sotto di un'etichetta da stampare. Includeremo anche una descrizione superficiale delle classi utilizzate e di come interagiscono tra di loro, con pseudocodice esplicativo annesso. Infine si concluderà con una panoramica sui risultati ottenuti su una base di dati di test generata ad hoc. Per il test su dati reali si guardi il capitolo 7.

### 6.1 *Scopi e strumenti di lavoro*

In CARGEN la visualizzazione di etichette per la toponomastica non era mai stata affrontata. Questo lavoro è il primo che introduce un approccio concreto con dei risultati visualizzabili nel famoso GIS OpenJUMP. Il progetto parte con l'obiettivo di realizzare una componente software in grado di generare una visualizzazione di etichette posizionate in modo intelligente, integrabile con i componenti già presenti sviluppati in precedenza all'interno del progetto per la generalizzazione, sia da 5000 a 25 000, che da 25000 a 50000.

Il lavoro è stato realizzato in Java, con l'ausilio di librerie Java già create all'interno del progetto CARGEN da Sandro Savino. Queste librerie rendono disponibili utili strumenti per il lavoro su tabelle e features, nonché su oggetti di tipo `Geometry`. E' stato inoltre necessario analizzare il codice sorgente di OpenJump (versione 1.4.0), per l'aggiunta di un plugin per le prove di visualizzazione. Le classi Java create infatti realizzano alcune interfacce proprie del core di OJ per permettere la visualizzazione grafica delle etichette con stili diversi.

Per quanto riguarda l'hardware, tutte le classi sono state create e testate su di un sistema portatile Sony Vaio VPCEB1C5E dotato di cpu dual core Intel i5-430M (4 core logici) @2,27GHz, gpu AMD RadeonHD5650M, ram DDR3 4GB dual channel a 1066MHz e HDD da 500GB a 5400rpm. Il sistema operativo della macchina è Windows 7 Home Premium a 64bit. Per lo sviluppo del codice è stato utilizzato l'IDE Eclipse, mentre per la visualizzazione e il test dei risultati è stato utilizzato il GIS OpenJump. I dati di partenza arrivavano dal DBT al 5000, ovvero sono stati sfruttati dati cartografici esistenti e disponibili all'interno del progetto CARGEN. Per test approfonditi sul funzionamento delle classi sono stati creati vari layer con geometrie inventate, simulando una mappa reale, per quanto possibile. Il problema, come evidenziato in precedenza, è che, anche se esistono ormai le direttive definitive per la creazione di database cartografici con dati standard per tutte le regioni d'Italia (dati da IntesaGIS<sup>1</sup>), con inclusa la trattazione completa dei dati della toponomastica, i dati utilizzati per il progetto non sono aggiornati e gli attributi specifici relativi alla toponomastica nei database in CARGEN sono assenti o largamente incompleti.

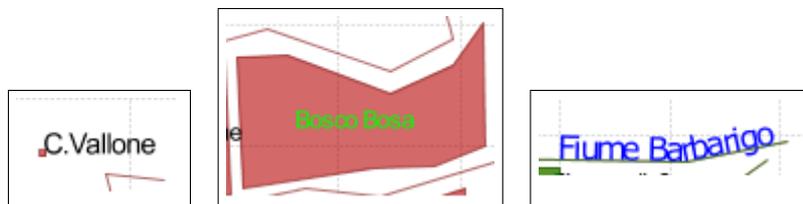
I dati sfruttati venivano caricati in RAM e processati direttamente, senza l'utilizzo di componenti e librerie per la comunicazione client-server necessarie in precedenza per altri lavori in CARGEN. Questo ha permesso una maggior libertà di modifica sui dati stessi e la possibilità di effettuare test in modo immediato sulla macchina client, senza l'utilizzo di intermediari per il caricamento dei dati da server. Ovviamente questo processo di lavoro è molto più snello e rapido.

## *6.2 Le etichette*

La prima parte del mio lavoro è stata votata alla ricerca di un metodo per stampare etichette sul software OpenJump, in modo poi da poter testare più facilmente i successivi algoritmi di label placing, appoggiato da una visualizzazione grafica ad hoc. Visto che le etichette vanno ad associarsi ad elementi cartografici, c'era la necessità di doverle creare adattabili e che si potessero scrivere con diverse angolazioni e su linee spezzate (si pensi ad esempio al nome di un fiume il cui toponimo segue il suo corso sulla mappa). Il software GIS OpenJump permette la visualizzazione di etichette associate ad ogni feature dei layer caricati, ma permette solo la stampa di scritte orizzontali, eventualmente con una certa angolazione definita da un attributo

---

<sup>1</sup> IntesaGIS è un'unione concordata in ambito nazionale italiano per l'accordo su direttive per la creazione di database topografici di interesse generale. Per info consultare [http://www.centrointerregionale-gis.it/script/Documenti\\_intesa.asp](http://www.centrointerregionale-gis.it/script/Documenti_intesa.asp)



**Fig. 6.1:** Le tre immagini rappresentano le tre possibilità di stampa delle etichette in base ai tre rispettivi elementi geometrici. L’etichetta puntuale viene stampata in alto a destra rispetto al punto, l’etichetta areale viene stampata nel punto centrale (centroide) dell’area e l’etichetta su linea spezzata viene stampata seguendo l’andamento della spezzata stessa.

interno della feature. Questo approccio risultava limitante per il progetto, così è stata creata da zero una classe per la generazione di etichette associate ad una linea spezzata (polyline).

Per comprendere poi come effettivamente viene creata l’etichetta per ogni feature bisogna prima specificare come si è deciso di classificare gli oggetti topografici. Questa distinzione viene fatta sull’attributo geometrico associato ad ogni feature. In ogni database cartografico, ormai, gli elementi vengono associati a tre tipi di geometrie: punto, area (poligono) e linea spezzata. Per il labeling, è stata utilizzata la stessa divisione, perciò gli oggetti vengono distinti in:

- *punti*: che possono rappresentare comuni, cime montuose, punti quotati, ecc.
- *linee spezzate*: rappresentano strade, fiumi e corsi d’acqua di vario tipo.
- *aree*: laghi, regioni, confini territoriali amministrativi, boschi, ecc.

Al variare della scala variano anche i tipi di geometria associati ad elementi comuni. Ad esempio alla scala 1:50000 un fiume sarà rappresentato da una linea, mentre alla scala 1:5000 da un’area (poligono). In Fig. 6.1 si può notare come vengano stampate le etichette nei tre diversi casi. Tutte le etichette possono essere stampate in font, grandezze e colori diversi.

Ultimo ma non meno importante, le etichette possono muoversi in diverse posizioni rispetto al loro elemento geometrico di riferimento. Ognuna di esse può inoltre valutare il suo grado di leggibilità (che sarà chiamato “felicità”) autonomamente in base alla posizione in cui si trova. Questo risulta fondamentale per riuscire a spostare le etichette in modo intelligente sulla mappa, come si vedrà successivamente.



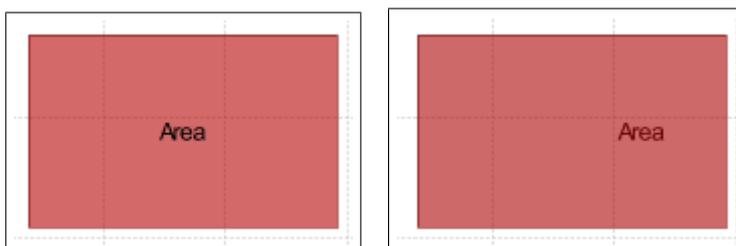
**Fig. 6.2:** Le tre immagini rappresentano tre possibili posizioni di stampa delle etichette puntuali. La prima è la migliore, la terza la peggiore, cioè con un minor grado di “felicità”.

Riassumendo, l’etichetta è un oggetto singolo associato ad un elemento geometrico che ha funzionalità di stampa in diversi formati e colori e di movimento all’interno di un’area limitata definita dall’oggetto geometrico di riferimento, sia esso un punto, una linea spezzata o un’area (poligono).

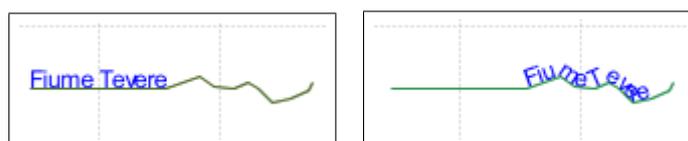
### 6.2.1 Qualità di un’etichetta in base alla posizione rispetto al suo riferimento

Ogni etichetta porta con sé un parametro che rispecchia la sua “felicità”, ovvero il grado di soddisfazione in base alla sua posizione rispetto all’elemento a cui è associata. Più l’etichetta è leggibile e più sarà felice. Questo parametro valuta esclusivamente la posizione rispetto all’elemento geometrico associato, senza tener conto di altri elementi geometrici o etichette sulla mappa. Ognuno dei tre tipi di etichetta ha un diverso metodo per calcolare la sua “felicità”. Nel processare questi valori ci si è attenuti, per quanto possibile, alle direttive e ai consigli elencati nel documento dell’IGM sulla toponomastica [vedi 7, pp. 13-28]. Queste delineano le regole del buon cartografo. Vediamo brevemente come vengono calcolati questi valori:

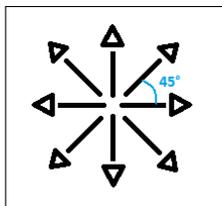
- *Etichetta puntuale*: l’etichetta rispetto al punto si può posizionare solo in quattro zone diverse, ovvero in alto a destra, in basso a destra, in basso a sinistra e in alto a sinistra. La posizione dove l’etichetta risulta più leggibile è la prima, in alto a destra rispetto al punto. Ecco quindi che la felicità massima si ha quando il toponimo viene visualizzato in questa posizione. La posizione con minor felicità è quella in alto a sinistra (vedi Fig. 6.2).
- *Etichetta areale*: l’etichetta ha maggior felicità quando viene stampata al centro dell’area, e avrà via via minor felicità nell’allontanarsi da questo punto ideale (vedi Fig. 6.3).
- *Etichetta su linea spezzata*: l’etichetta ha maggior felicità quando viene stampata in un tratto poco segmentato. Più la linea è fatta a zigzag meno felice è l’etichetta. (vedi Fig. 6.6).



**Fig. 6.3:** Le due immagini rappresentano due possibili posizioni di stampa delle etichette areali. La prima è la migliore, la seconda la peggiore, cioè con un minor grado di “felicità”. Infatti è posizionata più lontana dal centro.



**Fig. 6.4:** Le due immagini rappresentano due possibili posizioni di stampa delle etichette su linea spezzata. La prima è la migliore, è su posizione ideale, su linea retta, la seconda la peggiore, infatti è stampata su di un tratto molto segmentato.



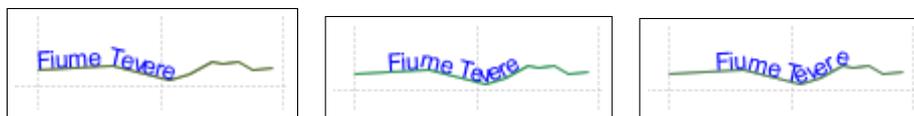
**Fig. 6.5:** L'etichetta di un'area si sposta della metà della sua larghezza ed altezza in una delle otto direzioni disegnate in figura.

Per la scelta delle varie posizioni sono state seguite le linee guida dettate da Imhof nel 1975 in un saggio riassuntivo [vedi 8] sul *label placing*. Il grado di felicità di ogni etichetta è valutato autonomamente al variare della sua posizione. Ma come si può cambiare il posizionamento di un'etichetta? Quali spostamenti vengono compiuti e con che logica?

### 6.2.2 Come si spostano le etichette

Ogni etichetta, quando viene creata, viene posizionata nella sua posizione iniziale, che spesso coincide con quella a maggior grado di felicità. Allo stesso tempo, ogni etichetta può essere spostata su altre posizioni predefinite. Infatti, esiste una funzione, chiamata *move*, che se invocata, agisce sull'etichetta spostandola di “un passo” in avanti. Questo passo è diverso per ognuno dei tre tipi di etichetta:

- *Etichetta puntuale*: Ad ogni chiamata di *move* l'etichetta si sposta in senso orario. Esistono solo quattro posizioni diverse possibili per il posizionamento rispetto al punto (vedi Fig. 6.2). Quando l'etichetta è posizionata in alto a sinistra, cioè nell'ultima posizione disponibile, la successiva chiamata di *move* la riporta in posizione iniziale, ovvero in alto a destra rispetto al punto.
- *Etichetta areale*: quando l'etichetta viene creata viene stampata al centro dell'area, ovvero il centro dell'etichetta sarà il centro dell'area stessa. Alla chiamata di *move*, questa viene spostata casualmente verso una delle otto direzioni prestabilite, come si può notare in Fig. 6.5. Se l'etichetta dovesse toccare i bordi dell'area, l'algoritmo di spostamento cambierà la direzione di spostamento, scegliendone una a caso tra le altre sette disponibili (non sceglierà mai l'inversa della direzione appena intrapresa, per evitare di far tornare l'etichetta sulle stesse posizioni).



**Fig. 6.6:** Le tre immagini mostrano, da sinistra a destra, la posizione originaria dell'etichetta su spezzata seguita dai due successivi spostamenti generati da due chiamate successive dell'algoritmo *move*.

- *Etichetta su linea spezzata:* l'etichetta ha maggior felicità quando viene stampata in un tratto poco segmentato. Più la linea è fatta a zigzag meno felice è l'etichetta (vedi Fig. 6.6). Quando viene chiamato *move* l'etichetta si sposta lungo la linea associata muovendosi di  $1/4$  della sua lunghezza (lunghezza del testo). La direzione di spostamento è verso destra. Quando l'etichetta raggiunge la fine della linea, ritorna alla posizione originale di partenza.

La necessità dell'implementazione degli spostamenti delle etichette deriva dal bisogno di posizionare le stesse in diversi modi per poter rendere la mappa più leggibile, cercando di evitare sovrapposizioni e situazioni di grande affollamento dovute all'alta densità di elementi da disegnare in punti critici della mappa.

### 6.3 Valutazione densità di elementi sottostante alle etichette sulla mappa

Come già precedentemente discusso, il problema del *label placing* su mappe deve considerare non solo le sovrapposizioni tra etichette, ma deve anche gestire l'affollamento esistente al di sotto delle etichette stesse. Per comprendere il problema si veda Fig. 6.7.

In questa sezione si andrà a presentare il metodo utilizzato nel mio lavoro per dare una valutazione del problema per applicarlo poi in pratica nella scelta del posizionamento delle etichette. L'approccio è basato su di un semplice algoritmo che calcola la densità di una mappa in funzione del numero di vertici delle geometrie presenti sulla stessa.

#### 6.3.1 La matrice di densità

Ogni mappa geografica consiste in un insieme di elementi geometrici. Ogni geometria, sia essa un punto, una linea o un poligono (anche con buchi

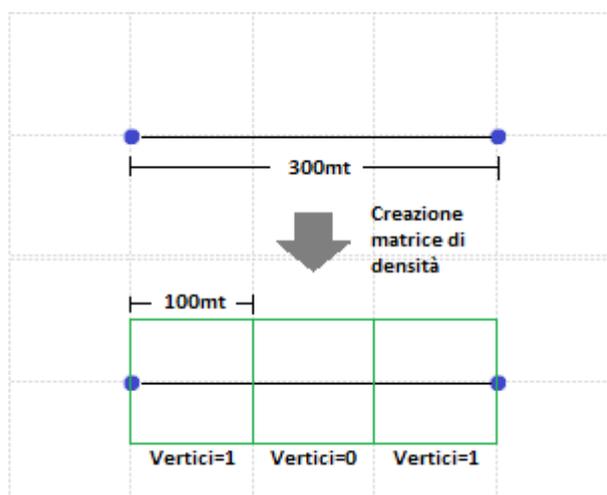


**Fig. 6.7:** In questo esempio si può notare come l'etichetta sia poco leggibile in quanto l'area in cui viene stampata è sovrappollata. Sarebbe opportuno spostarla in una zona della mappa più libera.

all'interno) è, in sostanza, un insieme di vertici più o meno ordinati. L'idea è quella di valutare l'affollamento di un'area in base al numero di vertici presente all'interno della stessa. Per fare questo è stata creata una matrice che divide una zona in quadrati di lato arbitrario, dove ogni quadrato tiene conto del numero di vertici presenti all'interno del proprio perimetro. In questo modo si ha un parametro qualitativo che rispecchia la reale densità della mappa.

Per rendere omogenea la valutazione è stato creato un algoritmo per l'aggiunta di vertici a elementi geometrici come linee ed aree, in quanto possono coprire una vasta area e quindi contribuire all'affollamento della mappa, senza che i vertici effettivamente presenti vengano rilevati. Per comprendere il problema e la soluzione adottata, viene presentato un esempio esplicativo (vedi anche Fig. 6.8). Si supponga di dover valutare il contributo di densità di una linea composta da due vertici, che copre un'area di 300 metri in tutta la sua lunghezza. Per la matrice di densità si decida di utilizzare un passo di campionamento (lunghezza del lato del quadrato) di 100 metri. Se si andasse a contare il numero di vertici presente la matrice risultante sarebbe divisa in 3 quadrati, dove nel primo e nell'ultimo saranno contati un vertice, e in quello centrale nessuno. Questa valutazione, però, non coglie con precisione sufficiente la realtà. Infatti, anche se non è fisicamente presente alcun vertice nella lunghezza della linea, questa crea ingombro all'interno della mappa, e, al piazzamento di un'etichetta in quella zona, creerebbe difficoltà di lettura. Ecco quindi la necessità di aggiunta di vertici alla geometria per permettere all'algoritmo di contagio dei vertici un risultato migliore e più vicino alla realtà.

Per ovviare al problema è stato implementato un algoritmo di addensamento (vedi Fig. 6.9) di geometrie che funziona in modo ricorsivo: riceve la geometria e per ogni coppia di vertici aggiunge un vertice a metà della distanza tra i due finché la distanza non risulta minore del passo di campionamento scelto per la matrice di densità. In questo modo ogni quadrato della matrice



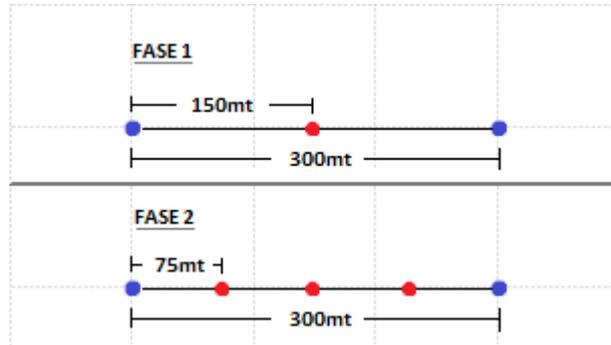
**Fig. 6.8:** Con una linea di 300mt si decide di valutare la densità di vertici suddividendo l'area in quadrati di  $100 \times 100 \text{mt}^2$ . Si nota come la matrice, in questo modo, non riesca a considerare efficacemente la densità di vertici su tutta la lunghezza della linea.

attraversato da una linea di una geometria avrà conterà almeno un vertice all'interno del suo perimetro (vedi Fig. 6.10).

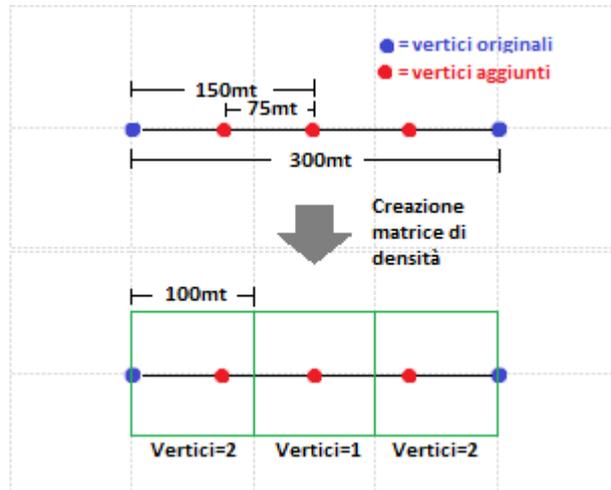
La soluzione combinata dell'algoritmo di densificazione delle geometrie e della creazione della matrice di densità dà uno strumento concreto per la valutazione dell'affollamento di elementi geometrici all'interno di una mappa. La generazione della matrice è inoltre molto semplice e veloce, requisito indispensabile quando si processano layer con centinaia di migliaia di features con milioni di vertici geometrici potenzialmente disponibili. La creazione della matrice viene calcolata in tempo lineare rispetto al numero di geometrie presenti nel layer preso in considerazione come mappa. Per comprendere "come" risulti agli occhi della matrice una mappa con alcuni elementi geometrici si può osservare Fig. 6.11, ovvero la mappa viene semplificata ad un insieme di vertici.

La soluzione è efficace nel rappresentare l'effettivo affollamento della mappa, come si può notare in Fig. 6.12, dove è stato assegnato un colore più saturo ai quadrati con all'interno un maggior numero di vertici. Ad una rapida occhiata si vede come le zone più "affollate" della mappa originaria risultino ben chiare anche nella matrice di densità associata.

La matrice di densità sarà utilizzata all'interno dell'algoritmo di posizionamento delle etichette per evitare di stampare etichette in zone della mappa troppo dense di vertici, dove potrebbero risultare di difficile lettura o



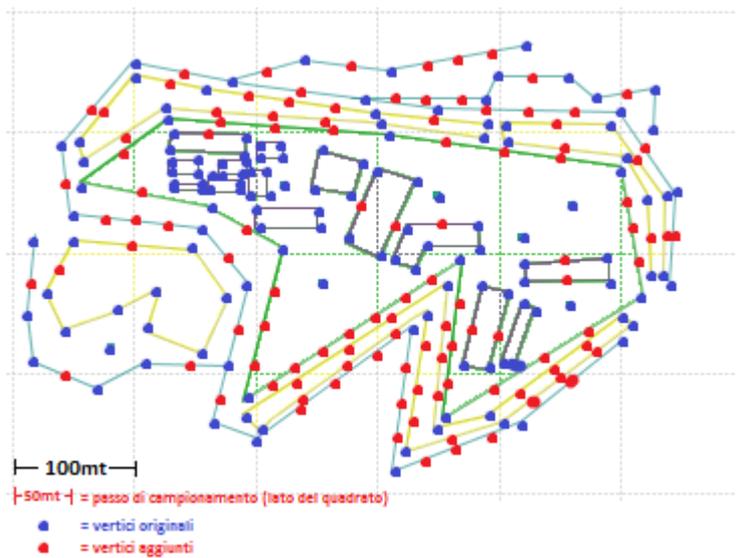
**Fig. 6.9:** Come si può notare in figura dove vengono illustrate le varie fasi, l’algoritmo di densificazione aggiunge vertici alla geometria ricorsivamente.



**Fig. 6.10:** Prima di creare la matrice di densità si aggiungono vertici alla geometria come descritto in precedenza. La matrice sarà diversa e darà una rappresentazione molto più omogenea dell’affollamento di vertici della mappa.

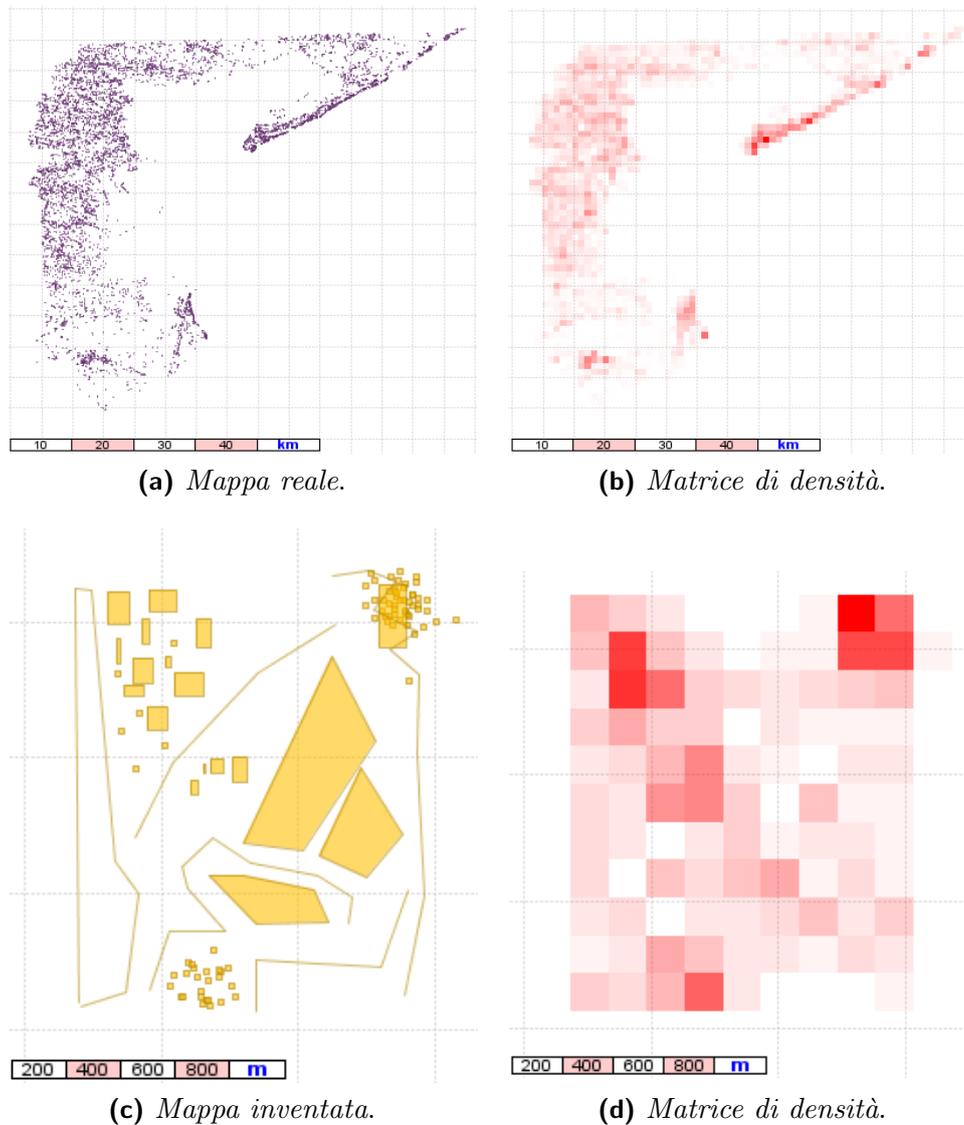


(a) *Mappa reale.*

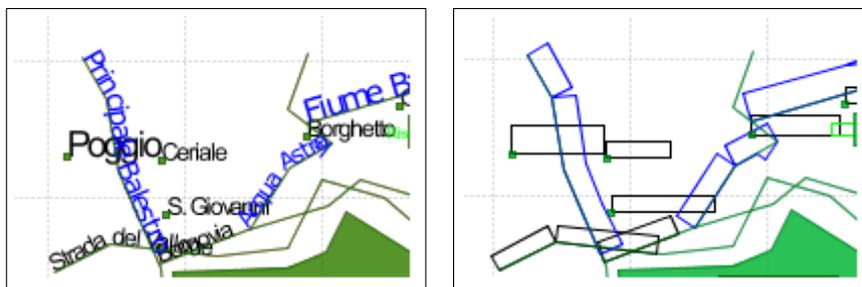


(b) *Vista dalla matrice.*

**Fig. 6.11:** Si osservi come la mappa appaia realmente, con tutti gli elementi geometrici, nella prima figura, e come venga “vista” dalla matrice, ovvero come un insieme di vertici, alcuni dei quali sono stati aggiunti dall’algoritmo di densificazione delle geometrie.



**Fig. 6.12:** Nelle due immagini in alto si nota un layer basato su dati reali con la rispettiva visualizzazione della matrice di densità con passo di campionamento di 1000mt. Nelle due immagini in basso la matrice di densità è generata su di un layer fittizio, con elementi inventati che simulano una mappa. In questo caso il passo di campionamento della matrice è di 150mt. In entrambi i casi si può apprezzare la qualità della matrice che riesce a conservare l'affollamento originale delle mappe.



**Fig. 6.13:** Nell'immagine si nota la differenza tra come venga visualizzata l'etichetta e quali siano i suoi bordi, con i quali l'algoritmo di posizionamento trova le sovrapposizioni.

completamente illeggibili.

#### 6.4 *L'algoritmo di piazzamento e il concetto di priorità*

Finora si è visto come vengono rappresentate e spostate le etichette, in più si è visto come viene valutato l'affollamento della mappa dove saranno disegnati i toponimi. Ora che sono stati presentati questi mezzi, si può spiegare come vengono sfruttati dall'algoritmo di posizionamento per generare una soluzione di stampa applicata ad un layer. Si andrà a vedere su quali concetti si basa l'algoritmo per la scelta della priorità di stampa delle etichette, come queste vengano spostate e come venga valutata la qualità della "configurazione" finale di etichette stampate. In sostanza si analizzeranno pezzo per pezzo i passi fondamentali compiuti dall'algoritmo per generare una soluzione accettabile.

##### 6.4.1 *Come vengono individuate le sovrapposizioni tra etichette*

Ogni etichetta è associata ad una stringa di testo e ad una geometria. Per individuare le sovrapposizioni l'algoritmo di posizionamento sfrutta una funzione dell'etichetta che ne calcola l'ingombro reale sulla mappa e, in sostanza, restituisce una geometria che rappresenta l'area occupata dal testo dell'etichetta (vedi Fig. 6.13).

##### 6.4.2 *Le priorità delle etichette*

L'algoritmo di posizionamento analizza le collisioni tra tutte le coppie di etichette, ma per decidere come spostarle, quali spostare, quali saranno

disegnate e quali eliminate, si basa su di un parametro associato ad ogni etichetta, ovvero un indice di priorità.

Ogni etichetta ha un parametro, che non è altro che un numero intero che va da 1 a 4, che ne indica la priorità di stampa. Dove 1 significa priorità massima (il toponimo deve essere visibile sulla mappa), e 4 significa priorità bassa (se collide con altre etichette sarà eliminata). Quattro valori di priorità sono sufficienti per ottenere una classificazione abbastanza precisa di tutti i toponimi da stampare sulla mappa. E' da sottolineare come l'assegnazione della priorità debba avvenire prima della generazione dell'etichetta, ovvero il parametro dovrebbe già essere presente all'interno delle tabelle del database conetnenti le features della mappa. La priorità non è autovalutata dall'algoritmo nè dall'etichetta stessa. E' un valore che andrebbe assegnato a priori tramite una valutazione logica e ponderata sull'effettiva importanza del toponimo.

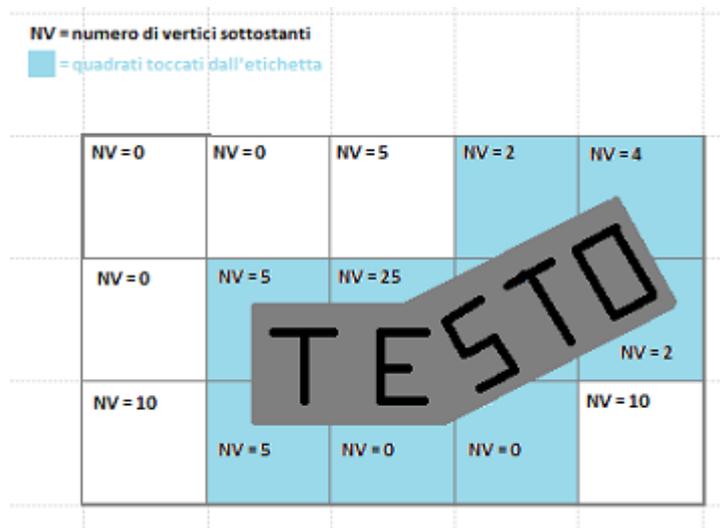
#### 6.4.3 Generazione matrice di densità e tabella etichette

L'algoritmo di posizionamento agisce come filtro su di una tabella di etichette precedentemente creata. In sostanza, da un grande insieme di toponimi, decide quali disegnare poi sulla mappa e dove disegnarli. Ma da cosa viene creata questa tabella? In sostanza si riceve in input un **Layer**, o più di uno, cioè un insieme di **Feature**, i dati cartografici. Ogni **Feature** contiene un elemento geometrico, detto **Geometry**. E' su questi dati che vengono generate le due strutture dati principali su cui agirà poi l'algoritmo di label placing:

- *La tabella delle etichette* per ogni **Feature** viene creata un'etichetta, **Label**, associata all'attributo **Text** che contiene una stringa di testo che rappresenta il toponimo. Tutte le etichette vengono salvate in una tabella.
- *La matrice di densità:* sui **Layer** ricevuti, si genera una matrice di densità come precedentemente descritto, con un passo di campionamento di 50 metri, ovvero la lunghezza del lato del quadrato della matrice. Questo valore è stato scelto perché in media un'etichetta al 25 000 occupa un'area di circa 400x50mt<sup>2</sup> sulla mappa. Questo parametro può essere più rilassato nel caso si lavori con mappe a scala minore.

#### 6.4.4 La felicità dell'etichetta singola e quella generale

In precedenza è stato descritto come ogni etichetta valuti la qualità del suo posizionamento rispetto alla sua geometria in modo automatico. L'algoritmo



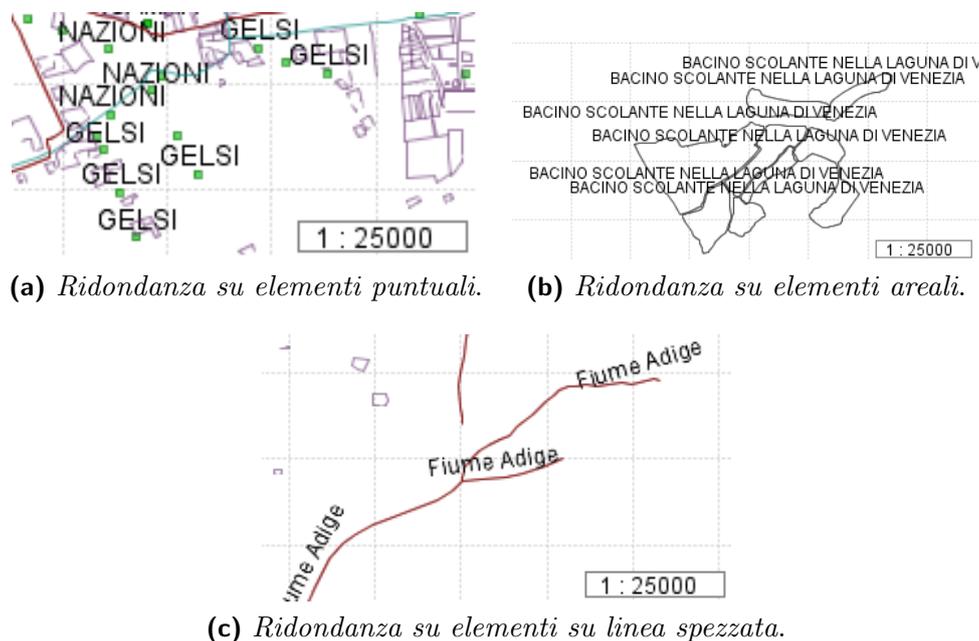
**Fig. 6.14:** Si osserva un esempio esplicativo del risultato ottenuto alla funzione propria della matrice di densità che restituisce il numero di vertici totale presente all'interno dei quadrati toccati dall'area dell'etichetta. Contando il numero di vertici sui quadrati azzurri si ottiene un totale di 43.

di posizionamento, però, aggiunge un ulteriore parametro per la valutazione della qualità, ovvero l'affollamento sottostante l'etichetta.

Per fare questo sfrutta una funzione della matrice di densità che restituisce il numero di vertici presenti al di sotto dei quadrati della matrice toccati dal bordo dell'etichetta presa in considerazione (vedi Fig. 6.14).

In questo modo si può valutare la qualità del posizionamento rispetto a due parametri:

1. *posizione rispetto alla geometria associata:* più è leggibile e maggiore è il valore di felicità assegnato (varia da 0 a 1). Si ricorda che ogni tipo di etichetta utilizza un diverso algoritmo di valutazione di qualità.
2. *grado di affollamento sottostante l'etichetta:* anche in questo caso, più è leggibile l'etichetta e maggiore è il valore di felicità assegnato (da 0 a 1). Facendo un po' di prove si nota che quando ci sono meno di una decina di vertici sotto l'etichetta questa risulta facilmente leggibile, perciò la funzione assegna un parametro di felicità massimo quando i vertici sottostanti sono meno di 11 e via via minore fino ad arrivare ad un valore minimo quando il numero di vertici supera le 50 unità.



**Fig. 6.15:** Nelle immagini nsì notano alcuni esempi di etichette ridondanti, ripetute in zone ravvicinate, con creazione di sovrapposizioni e affollamento eccessivo. Questo fenomeno si presenta sia per elementi puntuali sia per elementi su linee (fiumi e strade), sia su aree (catene montuose, valli, bacini d’acqua).

Questi due parametri vengono mediati in questo modo e determinano la felicità di una singola etichetta:

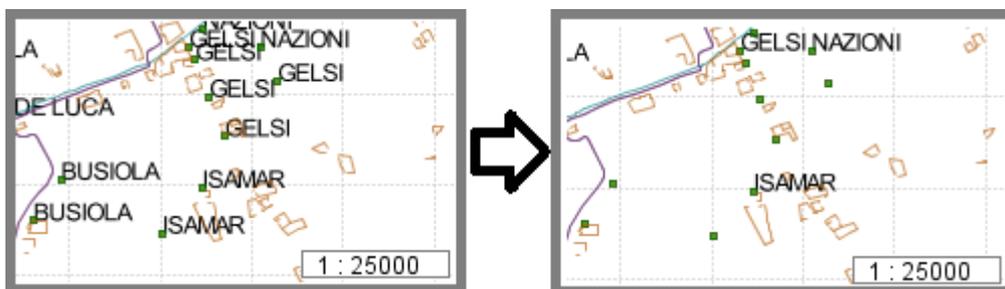
$$\text{Happiness} = \text{PositionHappiness} * 0.4 + \text{VertexHappiness} * 0.6$$

Esiste poi un parametro di felicità, chiamato **happiness** che mostra il grado di felicità dell’intera mappa sulle etichette. Non è altro che il valore medio tra tutti i valori di felicità di tutte le etichette disegnate.

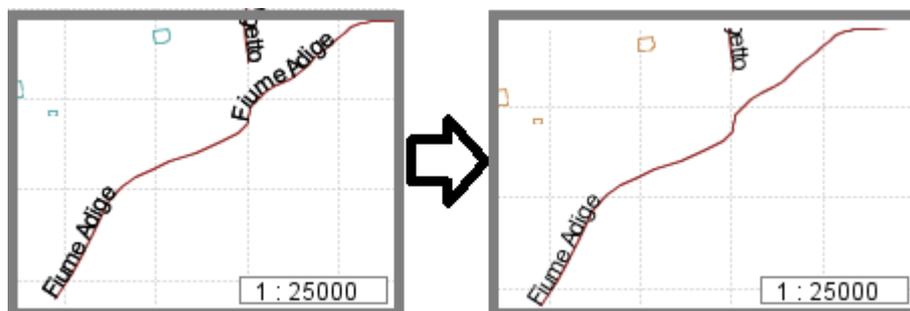
#### 6.4.5 Le ridondanze nelle etichette

Nei dati utilizzati esiste il problema della ridondanza di etichette. Ovvero, essendo ogni etichetta generata a partire dall’elemento geometrico associato, possono capitare casi in cui più elementi abbiano lo stesso toponomastico. Ecco che, alla stampa su mappa, nomi identici risultano ripetuti anche a minima distanza tra di loro, generando ripetizioni sgradevoli, come si può notare in Fig. 6.15.

Per tentare di risolvere il problema si è creato un catalogo delle etichette inserendolo in una mappa Hash chiave-valore, dove la chiave è rappresentata



(a) *Ridondanza risolta su elementi puntuali.*



(b) *Ridondanza risolta su elementi su linea spezzata.*

**Fig. 6.16:** Due particolari di una mappa in scala 1:25000 prima e dopo l'eliminazione delle ridondanze.

dal testo del toponimo e il valore è l'insieme delle etichette con associato tale toponimo. Alla creazione della tabella con tutti i dati delle etichette e ad ogni esecuzione dell'algoritmo `next` viene scansionato questo catalogo di etichette e vengono eliminate le etichette troppo "vicine" tra di loro, ovvero vengono eliminate le etichette con lo stesso toponimo che si trovano nel raggio di  $x$ km l'una dall'altra (questo parametro varia da scala a scala). Il risultato può essere apprezzato nel dettaglio in Fig. 6.16.

#### 6.4.6 I passi dell'algoritmo

Ora che sono state presentate le basi della struttura dell'algoritmo si andranno a vedere le fasi sequenziali dello stesso, tralasciando dettagli tecnici e concentrandosi sui punti principali. Ecco le varie fasi:

1. *Confronto coppia di etichette:* prende una coppia di etichette e confronta i bordi delle stesse per verificare se collidono.
2. *Gestione collisione:* se le etichette si sovrappongono decide quale etichetta spostare. La prima etichetta che si sposta è sempre quella a

priorità inferiore. Infatti ogni etichetta è piazzata nella sua posizione ideale, perciò si cerca di spostare quella meno importante. Dopo un certo numero di tentativi, se persiste la collisione, l'algoritmo sposterà anche l'etichetta a priorità maggiore.

3. *Collisione inevitabile*: se, dopo un determinato numero di spostamenti, le etichette rimangono sovrapposte, l'algoritmo considererà inevitabile la collisione e deciderà di non disegnare l'etichetta a priorità inferiore.
4. *Valutazione con matrice di densità*: quando le etichette sono state spostate e non collidono, l'algoritmo analizza la densità di vertici presente al di sotto delle stesse e, nel caso di etichette a bassa priorità, se queste dovessero trovarsi in zone troppo affollate, deciderà di non disegnarle.
5. *Ritorno al confronto*: ad ogni coppia di etichette sistemata ritorna al punto 1 e prende in considerazione un'altra coppia di etichette.

L'algoritmo agisce su tutte le etichette della tabella precedentemente generata e può essere fatto agire più di una volta. Ovvero, trovata una soluzione al termine di un primo passaggio, se lo si esegue un'altra volta, può trovare una soluzione diversa migliore, oppure può non eseguire alcun cambiamento se non trova collisioni o se la situazione non dovesse portare a miglioramenti.

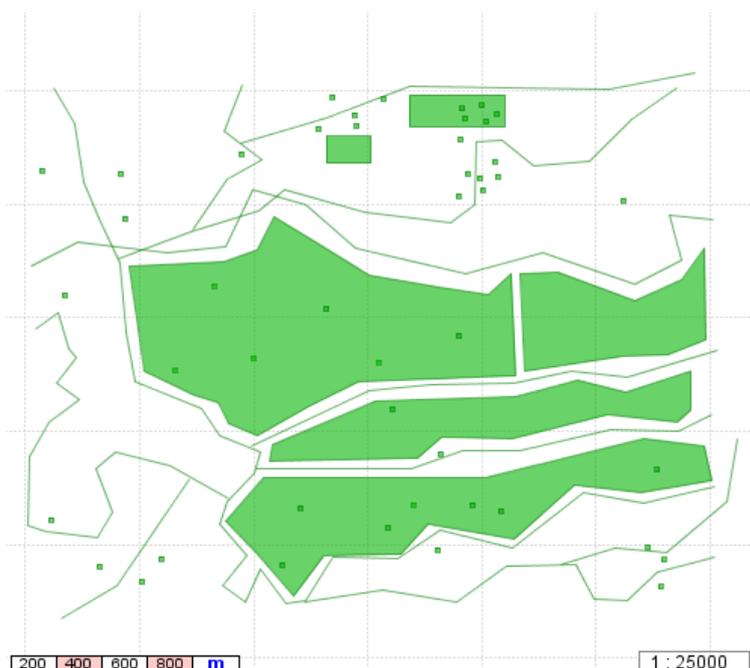
Al termine dell'esecuzione dell'algoritmo, viene effettuato un controllo sulla ridondanza delle etichette, eliminando etichette con identici toponimi stampate troppo vicine tra loro e, successivamente viene valutata la felicità generale della configurazione da disegnare, ovvero tutte le etichette da stampare, come descritto nella sezione precedente.

## 6.5 *Risultati e considerazioni*

In questa sezione saranno presentati il campo di test di esecuzione dell'algoritmo e i risultati ottenuti, considerando che non è un'applicazione reale ma i dati sono stati generati in laboratorio a scopo esclusivo di test.

### 6.5.1 *Il layer di test*

Per il test è stato creato un layer in formato shape popolandolo con una serie di features simulando una porzione di mappa 1:25 000. Nello specifico, il layer contiene 63 **Feature** ognuna con una **Geometry**, che può essere un poligono, una linea spezzata o un punto. Per una maggior comprensione, nella tabella è visualizzato un esempio di tabella del layer già popolata.



**Fig. 6.17:** Notiamo come il layer di test simuli una porzione digitale di una mappa in scala 1:25 000.

<i>TESTO</i>	<i>FONT TYPE</i>	<i>FONT SIZE</i>	<i>COLOR</i>	<i>PRIORITY</i>
S. Giorgio	Arial	12	Black	1
Borghetto	Verdana	10	Blue	3

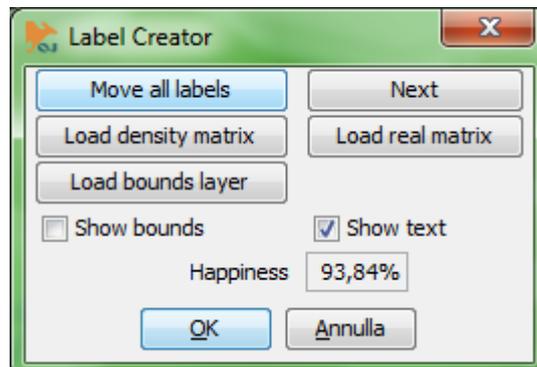
Con questi dati, diventa semplice generare le etichette per ogni **Feature** della tabella. Il layer di prova, visualizzato su OpenJump, si presenta come in Fig. 6.17.

### 6.5.2 *Interfaccia di test*

Per effettuare i test è stato creato un plugIn con semplice interfaccia grafica integrato in OpenJump. Questo plugIn genera un pannello grafico di gestione che permette l'attivazione di varie funzionalità, nonché la possibilità di avviare l'esecuzione dell'algoritmo di posizionamento sul layer visualizzato. Per i dettagli si veda Fig. 6.18.

### 6.5.3 *Risultati e confronti*

Per comprendere l'utilità dell'algoritmo di posizionamento è bene riassumere come vengono generate le etichette e come agisce l'algoritmo su di esse. Prima



**Fig. 6.18:** Per eseguire i test è stato utilizzato un plugIn con alcune funzionalità integrato in OpenJump.

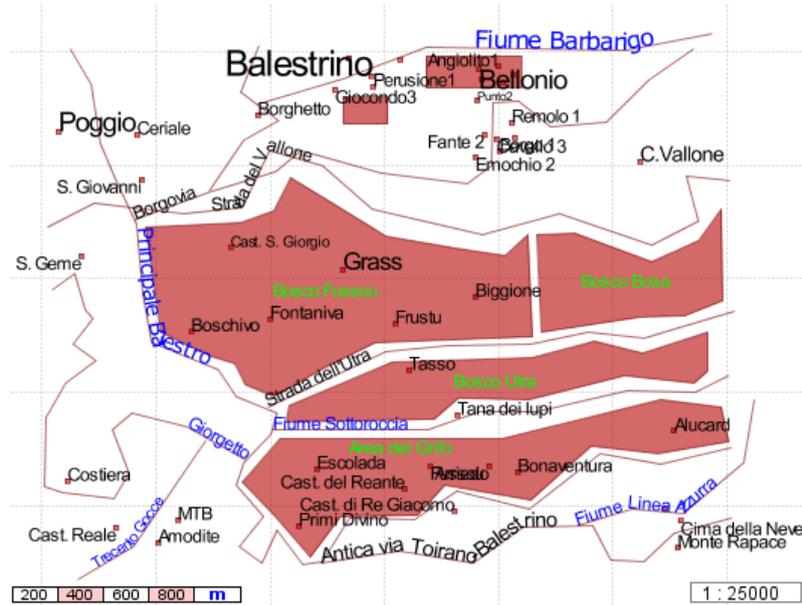
di far agire l'algoritmo viene generata un'etichetta per ogni **Feature** presente nella tabella e viene posizionata nella sua posizione iniziale, che varia in base alla geometria associata alla stessa, sia essa punto, linea o area (si veda la sezione 6.2). Poi si fa partire l'algoritmo di posizionamento. I risultati dopo l'esecuzione singola del processo possono essere apprezzati in Fig. 6.19.

L'algoritmo ha eliminato gran parte delle sovrapposizioni ma ha anche dovuto evitare di disegnare le etichette a più bassa priorità. In particolare delle 63 iniziali decide di stamparne sulla mappa 54.

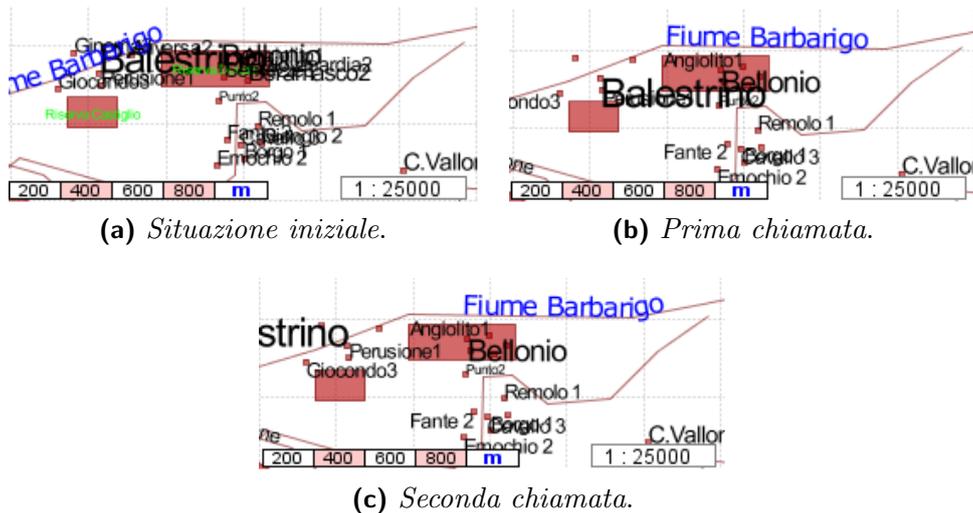
Un secondo passaggio dell'algoritmo agisce in modo meno sensibile sulla stampa delle etichette ma comunque rende leggermente migliore la disposizione sulla mappa, come si può notare in Fig. 6.20.

Per un'analisi al dettaglio, in Fig. 6.21 viene presentata la sequenza di tre screenshot su di una porzione di mappa che viene sensibilmente modificata al passaggio consecutivo dell'algoritmo.





**Fig. 6.20:** Questa è la mappa dopo il secondo passaggio dell'algoritmo sul layer di dati. Come si può notare la disposizione delle etichette è leggermente modificata.



**Fig. 6.21:** Come si può notare, dopo due chiamate dell'algoritmo di posizionamento, la leggibilità della porzione di mappa risulta migliorata, anche se persistono alcune collisioni.

## 7. TEST SU DATI REALI

Per avere un riscontro più veritiero sulle reali possibilità di applicazione dell'approccio proposto in sede di progetto, è stata eseguita una serie di test su dati reali, scaricabili dal portale dei Dati Territoriali della Regione Veneto<sup>1</sup>.

Nello specifico, sono stati utilizzati dati relativi alla zona della provincia di Chioggia, concentrandosi su di una singola sezione del DB25<sup>2</sup>. In questi dati, ogni elemento presente ha come attributo il suo toponimo, che sarà utilizzato per generare le etichette. I test sono stati eseguiti su tre layer di dati diversi:

- *Elementi idrici*: questi dati sono associati a geometrie composte da linee spezzate che rappresentano il corso dei fiumi e dei canali. Le etichette sono state stampate in colore blu.
- *Elementi stradali*: dati associati a geometrie composte da linee spezzate che rappresentano il corso di strade nazionali, regionali, provinciali e vicoli e vie minori. Le etichette sono state stampate in colore nero.
- *Località significative*: dati associati a geometrie puntuali che rappresentano una collezione di località significative, comprendendo nomi di comuni, frazioni, zone di villeggiatura, ospedali, ferrovie, piazze, ecc. Le etichette sono state stampate in colore nero.

In Fig. 7.1 si può notare come si presentino i dati, con e senza etichette applicate. Molti elementi non sono etichettati, proprio perché è stato eseguito un filtraggio e le etichette ridondanti non vengono disegnate. Su un totale di 2013 elementi, le etichette stampate sono 1060, circa la metà. Ciò significa che sono presenti molte ridondanze. Questo è un fenomeno inevitabile se, come si è deciso di fare in questo contesto, ogni elemento viene associato ad una etichetta.

L'esecuzione dell'algoritmo di posizionamento sui dati in esame produce dei risultati interessanti (vedi Fig. 7.2). Molte etichette vengono tolte dalla

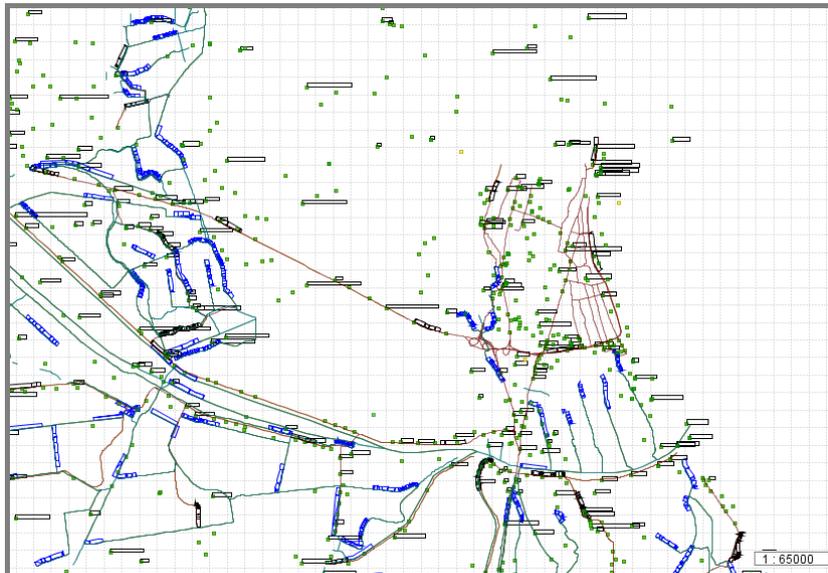
---

<sup>1</sup> Il materiale è scaricabile gratuitamente all'indirizzo <http://idt.regione.veneto.it/app/metacatalog/index?deflevel=1>

<sup>2</sup> Una sezione del DB25, in scala 1:25000, corrispondente ad un foglio delle carte IGM DB25, copre un'area territoriale di circa 150km<sup>2</sup>



(a) Sezione DB25 di Chioggia.



(b) Sezione con etichette applicate.

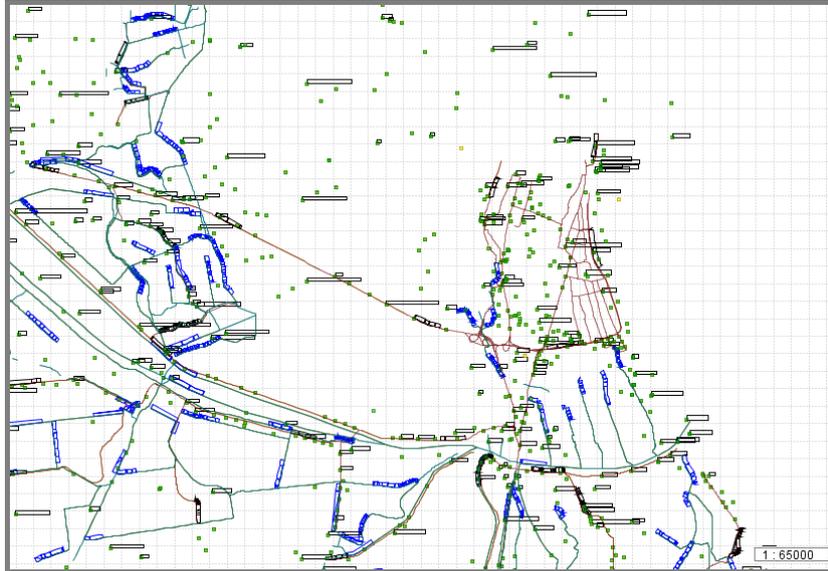
**Fig. 7.1:** Nelle immagini si nota la differenza tra i dati senza etichette applicate e i dati con etichette applicate. Su un totale di 2013 elementi ne vengono etichettati 1060 dopo l'eliminazione delle ridondanze.

---

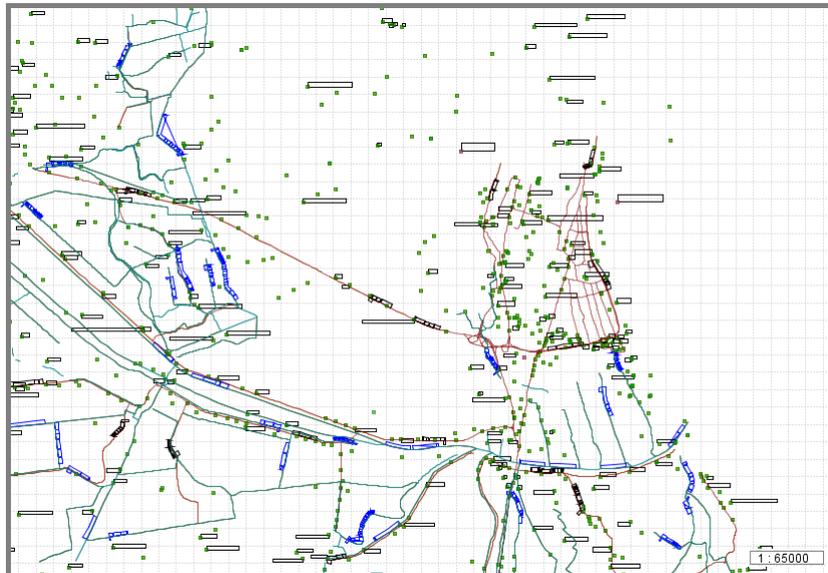
stampa. In totale ne vengono stampate 302 su un totale di 1060. Se si pensa che le regole IGM affermano che in una sezione del DB25, ovvero quella presa in considerazione, in media il numero di toponimi presenti dovrebbe essere pari a 300, si intuisce come questo risultato possa essere vicino alla realtà di una mappa concreta. L'algoritmo termina la sua esecuzione in 3 minuti e 32 secondi. La sua complessità temporale, a meno di costanti, è  $\Theta(n^2)$ , visto che confronta ogni etichetta con tutte le altre nella lista. I risultati non sono arbitrari, infatti nelle collisioni e negli spostamenti esiste un certo grado di casualità. Le etichette non si spostano sempre allo stesso modo e i confronti eseguiti non hanno un ordine preciso. Per questo si possono ottenere configurazioni diverse pur eseguendo il test sugli stessi dati di partenza.

Per un'analisi più dettagliata si osservino le immagini in Fig. 7.3, dove si può notare come l'algoritmo abbia agito su alcune zone della mappa.

L'algoritmo non riesce però a sfruttare tutte le sue caratteristiche, visto che nei dati utilizzati gli elementi non hanno diversi gradi di priorità. Tutte le etichette hanno perciò la stessa importanza e possono essere eliminati toponimi che, in una carta reale, non andrebbero eliminati.

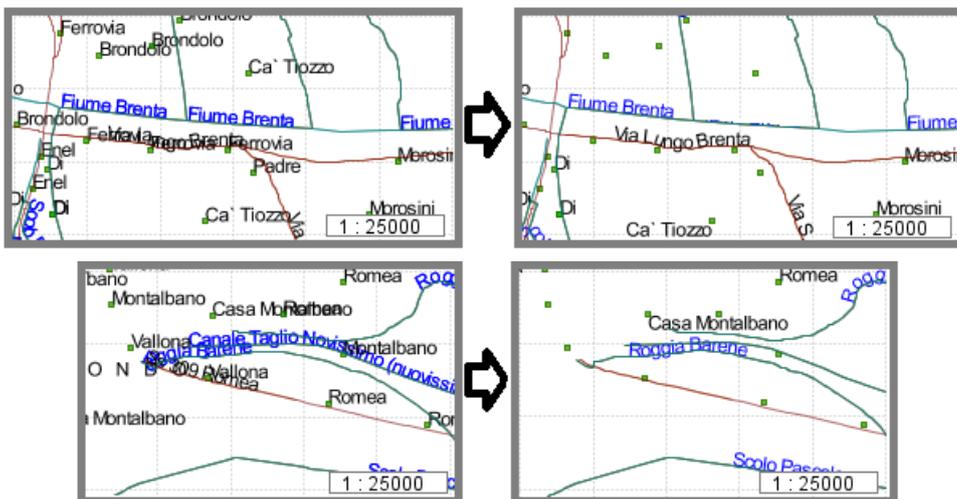


(a) *Prima del passaggio dell'algoritmo.*



(b) *Dopo il passaggio dell'algoritmo.*

**Fig. 7.2:** Le etichette stampate sono 302 su un totale di 1060. Dopo l'esecuzione dell'algoritmo si nota come molte collisioni siano state risolte e l'affollamento generale della mappa sia diminuito.



**Fig. 7.3:** Nelle immagini si nota il passaggio dal piazzamento di tutte le etichette (senza eliminazione ridondanze) alla soluzione data dall'algoritmo di posizionamento e dall'eliminazione delle ridondanze.

## 8. CONCLUSIONI E POSSIBILI SVILUPPI FUTURI

Questo lavoro di tesi si poneva l'obiettivo della trattazione del tema del posizionamento delle etichette della toponomastica all'interno del progetto CARGEN, in particolare si cercava di creare uno strumento concreto per la gestione della toponomastica nei database per la generazione di mappe in varie scale, seguendo le regole del DB25. Il lavoro si è diviso in due parti. La prima per trovare uno stile di visualizzazione e creazione delle label efficace e utilizzabile nel GIS OpenJump per permettere una valutazione visiva delle scritte sulle mappe, la seconda per trovare una possibile gestione di una struttura atta alla raccolta delle etichette e per tentare un approccio intelligente al piazzamento delle stesse sulle mappe. Si è arrivati a raccogliere i dati per ogni etichetta, generando un oggetto `Label` associato ad ogni `Feature` di una mappa digitale. Tutti gli oggetti `Label` vanno a popolare una tabella che verrà poi sfruttata dall'algoritmo di label placing nel valutare quali tra queste disegnaree quali non, aggiornando i dati relativi ad ogni etichetta.

Quindi è come se gli oggetti della toponomastica fossero trattati distintamente dagli oggetti geometrici del database ma sono ovviamente legati ad essi, perché ogni etichetta è associata ad un oggetto geometrico contenuto in una `Feature`. Un diverso approccio potrebbe essere stato quello di considerare un "layer della toponomastica", gestito indipendentemente, alla pari di ogni altro layer sulla mappa. Questo approccio rende però difficile l'associazione tra un toponimo e l'elemento cartografico di riferimento, visto che non esiste legame o riferimento presente nei dati tra i due. Si potrebbero gestire con meno vincoli di posizionamento le sovrapposizioni tra le etichette ma sarebbe arduo farlo in modo logico, considerando le giuste associazioni tra oggetti sulla mappa e toponimi di riferimento.

L'approccio scelto, allo stato attuale, genera delle ridondanze e un carico di informazioni non necessario nella creazione di una tabella di etichette. Ad esempio, nei database, un corso d'acqua non è costituito da una semplice geometria ma da molteplici geometrie, in relazione alla sua lunghezza, e la generazione di un'etichetta per ogni geometria risulta inutile, visto che il nome di un fiume non è stampato sulla mappa in modo così frequente. L'eliminazione delle ridondanze elimina in parte questo problema ma non è

---

sempore efficiente. Un altro esempio dove l'informazione del database non aiuta nella creazione delle etichette è quello dei confini amministrativi. Se per esempio si vuole stampare il nome di una regione, questa, nei dati, non è presente come unica geometria (un poligono) dove il suo perimetro definisce precisamente i suoi limiti, perciò non si può associare una sola etichetta che si possa spostare all'interno di un'area ben delimitata. D'altronde, associare ogni etichetta ad un oggetto geometrico, semplifica il controllo sui dati e permette di gestire il placing in modo più rapido, senza bisogno di interpretare il significato di un toponimo in relazione all'area circostante per identificare l'elemento di riferimento sulla mappa.

Per quanto riguarda il lavoro svolto con l'algoritmo di posizionamento, questo fornisce un approccio concreto, basato sulle priorità, alla soluzione del "label placing". Le basi su cui lavora l'algoritmo sono la priorità delle etichette e la felicità delle stesse in accordo al loro posizionamento sulla mappa e al grado di affollamento (densità di vertici) sottostante. L'approccio è sensato e produce risultati apprezzabili, ma ancora è in una fase di sviluppo embrionale. Infatti, al momento, non c'è possibilità di eseguire backtracking nelle posizioni delle etichette. Ovvero il grado di felicità delle etichette, sinonimo di qualità di posizionamento, non sempre va ad aumentare visto che non ha memoria dei suoi stati precedenti. Un possibile sviluppo futuro del lavoro potrebbe proprio essere l'affinamento dell'algoritmo con l'integrazione di strumenti di memorizzazione delle posizioni per poter garantire ogni volta scelte di posizionamento per le etichette realmente migliori.

## BIBLIOGRAFIA

- [1] John Ahn e Herbert Freeman. “A program for automatic name placement”. In: *Cartographica* 21.2–3 (1984), pp. 101–109.
- [2] Adriana C. F. Alvim e Éric D. Taillard. “POPMUSIC for the Point Feature Label Placement”. In: *Proc. Fifth Metaheuristics Internat. Conf. (MIC’05)*. Wien, 2005.
- [3] Lucas Bradstreet, Luigi Barone e Lyndon While. “Map-labelling with a multi-objective evolutionary algorithm”. In: *Proceedings Genetic and Evolutionary Computation Conf. (GECCO’05)*. New York, NY, USA: ACM Press, 2005, pp. 1937–1944. ISBN: 1-59593-010-8. DOI: <http://doi.acm.org/10.1145/1068009.1068335>.
- [4] Gildásio Lecchi Cravo, Glaydston Mattos Ribeiro e Luiz Antonio Nogueira Lorena. “A Greedy Randomized Adaptive Search Procedure for the Point-Feature Cartographic Label Placement”. In: *Comput. Geosci.* 34.4 (2007), pp. 373–386. DOI: <http://dx.doi.org/10.1016/j.cageo.2007.01.007>.
- [5] Shawn Edmondson et al. “A General Cartographic Labeling Algorithm”. In: *Cartographica* 33.4 (1997), pp. 13–23.
- [6] Olivier Ertz et al. “PAL - A cartographic labelling library”. In: *IICT-SYSIN, University of Applied Sciences, Switzerland* (2008).
- [7] IGM. *Specifiche IGM*. Giu. 2006.
- [8] Eduard Imhof. “Positioning Names on Maps”. In: *The American Cartographer* 2.2 (1975), pp. 128–144.
- [9] Claudia Iturriaga. “Map Labeling Problems”. Tesi di dott. School of Computer Science, University of Waterloo, 1999. URL: <http://www.cs.unb.ca/profs/citurria/Research/>.
- [10] Joe Marks e Stuart Shieber. “The Computational Complexity of Cartographic Label Placement”. In: *Harvard School Journal* (Febbraio 1993), p. 27.

- 
- [11] Hugo A. D. do Nascimento e Peter Eades. “User Hints for Map Labeling”. In: *J. Visual Languages & Comput.* 19.1 (2008), pp. 39–74. DOI: 10.1016/j.jvlc.2006.03.004. URL: <http://dx.doi.org/10.1016/j.jvlc.2006.03.004>.
- [12] Sandro Savino. “Il processo di generalizzazione cartografica: dalla Carta Tecnica Regionale al DB25 IGM”. Tesi di ricerca. Padova - Dipartimento DEI: Università degli Studi di Padova - Dipartimento di Ingegneria dell’Informazione, apr. 2007.
- [13] Alexander Wolff. “Automated Label Placement in Theory and Practice”. Tesi di dott. Freie University, mag. 1999.
- [14] Alexander Wolff et al. “A Simple and Efficient Algorithm for High-Quality Line Labeling”. In: *Taylor & Francis* (2000).