

Paolo Mogorovich

Sistemi Informativi Territoriali

Appunti dalle lezioni

L'informazione geografica

Cod.201 - Vers.E3B

- 1 - L'informazione geografica**
- 2 - Organizzazione dell'informazione in strati**
- 3 - La metainformazione**
- 4 - L'omogeneità delle classi**
- 5 - Georeferenziazione indiretta**

1 - L'informazione geografica

Consideriamo un'entità che ha una forte valenza territoriale, per es. una "strada"; la valenza territoriale di una strada è evidente, visto che il suo sviluppo sul territorio è una parte significativa della descrizione dell'entità stessa.

Se consideriamo una strada in modo "non spaziale", trascurando cioè il suo sviluppo sul territorio, possiamo descriverla tramite una serie di attributi: il suo nome, la sua lunghezza, la larghezza media, il costo annuo di manutenzione, il numero totale di incidenti, ecc. Tali dati possono essere organizzati in una tabella e sono costituiti da stringhe o da numeri. L'elaborazione di questi dati può portarci a conoscere grandezze derivate, come la superficie totale (approssimata), il numero di incidenti per chilometro, ecc.

Se consideriamo la strada in una logica "spaziale", le informazioni che ci viene naturale trattare sono di altro tipo; ad esempio se collega due centri abitati, quali altre strade incrocia, quali piazze attraversa, qual è il massimo dislivello, ecc. Rispetto all'approccio non spaziale, si trattano evidentemente informazioni di tipo diverso; tale differenza risalta ancora di più se osserviamo che, mentre le informazioni dell'approccio non spaziale descrivono l'oggetto "in quanto tale", quelle dell'approccio spaziale ci permettono di mettere in relazione l'oggetto con altri oggetti, sia dello stesso tipo che di tipi diversi. Le relazioni tra l'oggetto originario e altri oggetti sono in realtà il risultato di un'elaborazione che, per un osservatore che analizza una carta disegnata, avviene in modo naturale e intuitivo. La stessa elaborazione, all'interno di uno strumento informatico, sarà possibile perchè l'oggetto è appoggiato sul territorio e che, definito un sistema di riferimento, il posizionamento dell'oggetto, insieme a quello degli altri oggetti, è descritto tramite coordinate.

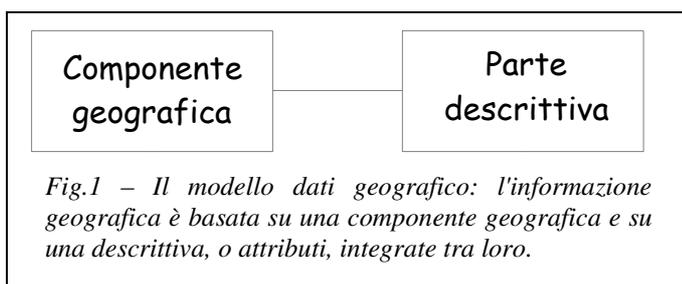
Quanto detto fa capire che gli oggetti che sono appoggiati sul territorio (e che quindi riusciamo a rappresentare su una mappa) sono caratterizzati dal fatto di possedere due tipi di informazione, diverse e integrate; un'informazione geografica e una descrittiva (gli attributi). Per esempio un edificio porta un'informazione geografica consistente nella sua posizione sul territorio e nella sua forma e una descrittiva consistente,

per esempio, nella sua tipologia, nel numero di persone che vi abitano, nel numero di piani; analogamente una strada porta un'informazione geografica consistente nel suo sviluppo sul territorio e una descrittiva consistente, per esempio, nel numero di corsie, nel tipo di asfalto, nel traffico.

L'informazione geografica si esprime tramite coordinate in un sistema di riferimento; alcune informazioni particolari, tipiche della maggior parte degli oggetti territoriali, quali la lunghezza (p.es. di una strada) e la superficie (p.es. di un edificio) sono attributi e non sono dati di tipo geografico, anche se sono ricavabili dalle coordinate.

L'informazione descrittiva di un oggetto territoriale può essere visualizzata scrivendo, a video o su carta, i numeri o le stringhe di caratteri di cui è composta; l'informazione geografica si visualizza correttamente solo tramite una rappresentazione cartografica.

La parte geografica e quella descrittiva sono concettualmente diverse; alcuni software memorizzano i due tipi di dati in due diversi DBMS, altri all'interno di un unico DBMS, ma anche in quest'ultimo caso la differenza concettuale tra i due tipi di informazione rimane e si evidenzia pensando alle diverse operazioni che possiamo applicare alle due componenti. Infatti, sulla parte descrittiva, che contiene "valori" attribuiti a caratteristiche dell'oggetto, possiamo applicare operatori di tipo tradizionale. Possiamo cioè relazionare oggetti diversi confrontando valori degli attributi: per esempio "se l'edificio E1 ha più vani dell'edificio E4", oppure "quante strade in quest'area sono non asfaltate". Sulla parte geografica, invece, possiamo applicare operatori diversi, quali la "vicinanza", la "adiacenza", la



“connessione”, ecc. e verificare, per esempio, se “Il Comune D1 confina col Comune R7” o se “la strada A1 è connessa con la strada A11”.

Gli operatori che agiscono sulla parte geografica possono essere di tipo geometrico, topologico o strettamente geografico; quelli che agiscono sulla parte descrittiva possono essere di tipo aritmetico, statistico, trattamento di stringhe, ecc.

2 – Organizzazione dell'informazione in strati

Gli oggetti di una mappa sono logicamente raggruppati in classi, ciascuna delle quali contiene al suo interno dati omogenei (la classe delle strade, quella dei ponti, quella delle formazioni geologiche, ecc.). All'interno di una classe sono contenuti più oggetti in genere ciascuno con una sua propria caratterizzazione in termini geografici (diversa posizione e forma) e descrittivi (p.es. tratto di strada statale o provinciale, oppure asfaltata o non asfaltata, ecc.). Le classi sono dette anche strati informativi o livelli (in inglese layer).

Una classe di oggetti è caratterizzata dal fatto di avere una specifica rappresentazione geometrica e una

struttura di attributi ben definita. Ad esempio una classe “strade” potrà essere definita da elementi lineari e da attributi che indicano il tipo e il nome della strada, il numero e la pavimentazione.

Ciascun istanza di strada sarà descritta da una linea sua caratteristica che la posiziona nello spazio e valori specifici per i quattro attributi; tuttavia tutte le istanze saranno descritte da linee e dai quattro attributi. Non ci potranno essere attributi diversi per una certa istanza e ciascuna istanza avrà tutti e quattro gli

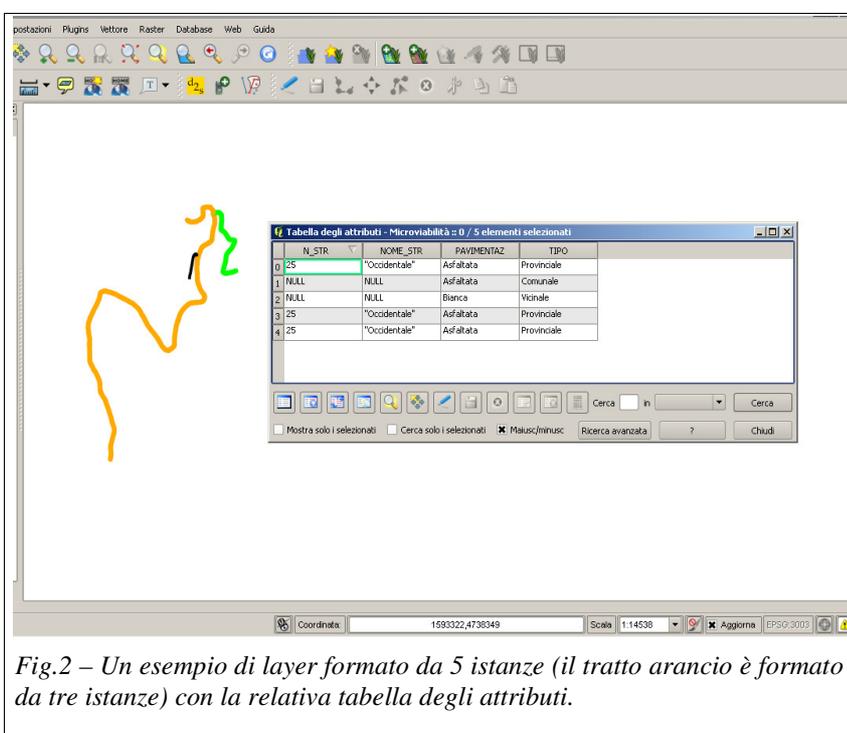


Fig.2 – Un esempio di layer formato da 5 istanze (il tratto arancio è formato da tre istanze) con la relativa tabella degli attributi.

attributi; eventualmente qualche valore potrà essere non conosciuto, e alla corrispondente casella sarà associata l'informazione “no data” o “NULL”, cioè “valore non conosciuto”.

Consideriamo oggetti appartenenti a livelli informativi diversi, p.e. strade ed edifici. Ciascuno di essi ha la medesima struttura generale, cioè la componente geografica integrata a quella descrittiva. La parte descrittiva è strutturata diversamente per le due classi: mentre la parte descrittiva della classe “strade” descriverà, ad esempio, il tipo di pavimentazione della strada, il suo nome, la larghezza, ecc., la parte descrittiva della classe “edifici” descriverà la destinazione d'uso dell'edificio, la sua altezza, il numero di vani presenti, ecc. Se la parte descrittiva è molto diversa per le due classi, al contrario la parte geografica è strutturata in modo molto simile; infatti, qualunque sia la classe, la descrizione geografica si basa su un unico criterio: le coordinate geografiche. Queste esprimono lo sviluppo di una casa in modo un po' diverso da come esprimono la curva di una strada, ma sono scritte nello stesso “linguaggio”, quello delle coordinate geografiche.

Questa caratteristica permette di applicare operatori spaziali a oggetti appartenenti a classi diverse; mentre in assenza di informazione spaziale non siamo in grado di mettere in relazione oggetti diversi, sfruttando il comune linguaggio con cui è codificata l'informazione spaziale possiamo confrontare, correlare, collegare, tramite operatori, oggetti eterogenei.

Gli operatori che agiscono su oggetti appartenenti a strati diversi sono la versione algoritmica di quello che facciamo quando, guardando una carta, mettiamo in relazione entità eterogenee. Tuttavia sulla carta i tipi di oggetti eterogenei e le loro caratteristiche sono limitati dalle caratteristiche della carta stessa, mentre nel mondo informatico non abbiamo, da questo punto di vista, alcun limite concettuale.

3 - La metainformazione

Abbiamo visto che l'informazione geografica è organizzata in classi o strati informativi e che all'interno di ciascuna classe sono presenti istanze di oggetti simili, ciascuna formata da una parte geografica e da una descrittiva. Esiste un altro tipo di informazione che riguarda ciascuna classe: tutti gli oggetti descritti all'interno di una classe possiedono alcune caratteristiche comuni. Per esempio per tutti gli oggetti di una certa classe le coordinate sono espresse in un certo sistema di riferimento, i dati si riferiscono alla stessa data, l'accuratezza geometrica è la stessa, l'Ente che li ha creati è lo stesso. È ovvio che queste ulteriori informazioni non vengono descritte nella tabella degli attributi, in quanto sarebbero "colonne" con lo stesso identico valore per tutte le istanze; tali valori sono memorizzati in un supporto a parte e vengono associati, com'è la loro natura, all'intera classe e non alle singole istanze. Se i dati relativi alle singole istanze sono l'informazione che noi abbiamo, questa ulteriore informazione è una "informazione dell'informazione" e come tale prende il nome di "metainformazione" o "metadati".

Poiché il primo accesso di un generico utente ad una classe di dati dovrebbe essere un'attenta lettura della metainformazione, è ovvio che sono state prese importanti iniziative per standardizzare la metainformazione, secondo strutture dati piuttosto complesse. Si rimanda alla letteratura specialistica per i necessari approfondimenti (vedi azioni CEN, ISO e Open Geospatial Consortium), e ci limitiamo qui a enfatizzare l'uso della metainformazione sia quando si utilizzano dati provenienti da fonti esterne (lettura), sia quando si costruiscono nuovi archivi (scrittura). Va anche detto che la presenza di metainformazione associata a classi di dati territoriali non è ancora molto diffusa, anche per una serie di difficoltà oggettive, come risalire a modalità di acquisizione per dati vecchi, per aggiornamenti di dati con caratteristiche diverse dall'archivio madre, ecc.

Qli attuali SW GIS gestiscono una certa parte di metadati quando trattano un layer: alcuni sono ricavati automaticamente dal layer e non sono di grande utilità per l'utente (numero di istanze, ampiezza geografica dell'area trattata, ecc.); ma sicuramente uno dei metadati di cui hanno bisogno è il sistema di riferimento adottato.

4 – L'omogeneità delle classi

Una classe, all'interno di un Data Base geografico, contiene un certo numero di oggetti aventi certe caratteristiche di omogeneità; questa omogeneità si esprime principalmente attraverso il tipo di primitiva, in genere unico per tutte gli oggetti, e gli attributi. Il tipo di primitiva e i tipi di attributi costituiscono il tipo di oggetto; all'interno del DataBase troveremo in quella classe tanti oggetti di quel tipo.

Supponiamo di definire una classe per descrivere gli edifici presenti sul territorio, dedicati ad uso abitativo. Le caratteristiche della classe potrebbero essere così descritte:

Parte geografica	Primitiva	AREA		
Parte descrittiva	Attributo 1	ID	Identificatore	Numero
Parte descrittiva	Attributo 2	VIA	Via su cui l'edificio ha l'accesso principale	Stringa
Parte descrittiva	Attributo 3	NC	Numero civico dell'accesso principale	Stringa
Parte descrittiva	Attributo 4	NAB	Numero delle persone che abitano nell'edificio	Numero

Parte descrittiva	Attributo 5	ASC	Indicatore della presenza di un ascensore	Binario
-------------------	-------------	-----	---	---------

Questa struttura descrive "l'oggetto tipo" e ciascun oggetto appartenente alla classe la rispetta nel senso che ogni istanza, cioè la descrizione di ogni singolo edificio, sarà formata da una primitiva geometrica areale, da un identificatore numerico, da una stringa che indica la via, ecc.

Nessun'istanza avrà una descrizione più ricca di quella definita, mentre, al contrario, non è detto che ciascuna istanza possieda valori per ciascun attributo; in pratica un attributo può essere obbligatorio o facoltativo e questo fa parte della descrizione della classe.

Per descrivere una classe potremmo usare una simbologia grafica molto semplice e intuitiva, che, nel caso considerato, è la seguente:



All'interno del Data Base Geografico dobbiamo immaginare poi di avere un numero indefinito di oggetti che condividono questa struttura, che possiamo rappresentare, in un esempio di 4 istanze, nel modo seguente:

AREA	-----	Id	VIA	NC	NAB	ASC
	-----	1	G.Bruno	1	12	si
	-----	2	G.Bruno	3	15	si
	-----	3	G.Bruno	5	4	no
	-----	4	G.Bruno	7	18	si

La definizione delle classi non è sempre semplice, in quanto non è sempre chiara una linea di demarcazione tra i diversi tipi di oggetti; ci possiamo trovare di fronte a oggetti che sono molto simili tra loro per certe caratteristiche, ma diversi per altre.

Proseguendo l'esempio precedente, supponiamo di aver a che fare, oltre agli edifici destinati ad uso abitativo, a altri edifici destinati ad attività industriale. Supponiamo ancora che gli elementi conoscitivi di nostro interesse per le abitazioni siano quelli descritti precedentemente, mentre, per quanto riguarda gli edifici industriali, interessi conoscere la Via, il Numero Civico, il numero di impiegati (NIMP) e la lavorazione prevalente (LAV).



Nel definire il modello della banca dati abbiamo due possibilità: o definire un'unica classe per ambedue i tipi di edifici o definire una classe per ciascun tipo di edificio. Il secondo caso non presenta alcuna complessità particolare; nel primo caso, invece, il nostro modello dati dovrà essere il seguente:



Questo modello dati implica, nel nostro caso particolare, quanto segue:

1. è necessario un campo "Tipo" per indicare se l'edificio ha destinazione abitativa o industriale (questo non sarebbe necessario nell'altra ipotesi); questo campo può assumere solo due valori: "abitativo" o "industriale"
2. in tutte le istanze in cui "Tipo" assume il valore "abitativo" i campi "NIMP" e "LAV" sono vuoti;
3. in tutte le istanze in cui "Tipo" assume il valore "industriale" i campi "NAB" e "ASC" sono vuoti.

Nel caso di 4 istanze di abitazione e 3 di fabbriche avremo una tabella degli attributi come in figura:

AREA	Id	VIA	NC	NAB	ASC	NIMP	LAV	TIPO
	1	G.Bruno	1	12	si			abitativo
	2	G.Bruno	3	15	si			abitativo
	3	G.Bruno	4			44	B01	industriale
	4	G.Bruno	5	4	no			abitativo
	5	G.Bruno	7	18	si			abitativo
	6	G.Bruno	8			16	B31	industriale
	7	G.Bruno	9			6	B01	industriale

In questo caso la fase di acquisizione dati richiederà qualche controllo in più e le query potrebbe risultare un po' più articolate; questi svantaggi sono compensati dal fatto di avere un'unica classe invece di due classi.

Il problema che abbiamo è pertanto quale strategia usare nel caso in cui si abbiano molte classi di oggetti lievementi differenti uno dall'altro: definire una classe per ciascun oggetto rende macchinosa e noiosa l'interazione con lo strumento informatico, mentre un accorpamento deciso rende eterogenea la classe. Nella scelta della strategia intervengono però altri fattori quali:

- l'esistenza o meno di un'unica fonte di dati per le varie classi elementari
- la diversa stabilità nel tempo delle varie classi elementari
- la diversa accuratezza delle informazioni
- ecc.

Di fatto non esiste una strategia definibile a priori, ma un'attenta analisi delle classi, le capacità del SW disponibile e soprattutto l'uso che si farà dei dati aiuteranno nella scelta ottimale. Nell'esempio proposto precedentemente, ad esempio, poiché il vantaggio di avere un'unica classe invece di due è modesto, la scelta migliore sarebbe di tenere le due classi separate. La questione, in generale, si presenta nel caso di abbondanza di informazione (molte classi e per ciascuna classe molti attributi). Nel caso in cui un insieme di classi abbiano pochi attributi, l'accorpamento è tecnicamente più semplice.

Nel caso dell'informazione geografica il problema dell'omogeneità delle classi si presenta in alcune situazioni tipiche:

- le strade che spaziano, per tipologia, dalle autostrade alle strade sterrate in mezzo ai campi
- i fiumi: da fossi asciutti in gran parte dell'anno a fiumi importanti
- infrastrutture per il trasporto di energia e per telecomunicazioni.

5 - Georeferenziazione indiretta

Abbiamo visto che la componente geografica dell'informazione territoriale si esprime tramite coordinate. Tuttavia esiste un caso in cui l'informazione geografica appare in modo indiretto, non tramite coordinate geografiche, ma tramite un riferimento ad altre entità spaziali. Quando si dice che una persona abita nel comune di Pisa, di fatto, la si localizza sul territorio, anche se con una bassa precisione. Il toponimo "Pisa" è un chiaro riferimento geografico di cui si suppone di conoscere le coordinate da un'altra fonte; in questo modo la localizzazione della persona avviene in modo indiretto. Casi tipici di georeferenziazione indiretta sono l'indirizzo (espresso da Città, Via e Numero Civico) e il Codice di Avviamento Postale.

La georeferenziazione indiretta espressa tramite Città, Via e Numero Civico (abbreviato in VNC) è di particolare importanza; infatti esiste una quantità enorme di oggetti/eventi che, per motivi amministrativi, possiedono tale tipo di georeferenziazione. Informazioni quali "dove una persona abita", "dove lavora", "dove è stato commesso un furto", "dove perviene un finanziamento", sono esempi di descrizioni di entità/eventi di cui si conosce una georeferenziazione indiretta.

La georeferenziazione indiretta si esprime tramite gli attributi di una tabella, come se fosse una parte descrittiva. Su questi attributi è possibile eseguire operazioni che sono un po' a cavallo tra operazioni tradizionali e operazioni descrittive (queste operazioni utilizzano operatori non geografici); per eseguire operazioni geometriche, invece, occorre trasformare la georeferenziazione indiretta in georeferenziazione diretta. Per questo è necessaria una mappa.

La tabella che segue riporta un semplice indirizzario, dove l'insieme di attributi Via+Numero_Civico+Città individua sul territorio un individuo, non tramite coordinate, bensì tramite un sistema di riferimento indiretto che può essere utilizzato direttamente da chi conosce la Città (mappa mentale), ma può essere anche trasformato in coordinate conoscendo una mappa reale della città.

Num. record	Cognome e nome	Via	Num. Civico	Città	Età
1	Pierini Giuliana	Ottone Rosai	3	Pisa	27
2	Archi Rita	della Repubblica	5	Pisa	35
3	Bargiacchi Eneo	Ottone Rosai	5	Pisa	28
4	Termia Emanuele	Ottone Rosai	17	Pisa	52
5	Francia Paolo	del Mercato	4	Pisa	45

Sui dati codificati in maniera indiretta si possono compiere operazioni non geografiche di ricerca tabellare, come ad esempio

“Quante persone hanno età compresa tra 20 e 30 anni ?”

oppure operazioni che possiamo definire “quasi geografiche” eseguibili tramite una ricerca tabellare; ad esempio:

“Quante persone abitano in Via Rosai ?”

mentre altre operazioni, anch'esse geografiche, richiedono di poter appoggiare su una mappa l'informazione VNC, trasformandola in coordinate, come ad esempio:

“Chi abita più vicino a Paolo Francia ?”

La traduzione da georeferenziazione indiretta in georeferenziazione diretta, cioè da VNC a coordinate, può avvenire in modo abbastanza rapido tramite tabelle di conversione, dette geovocabolari, che trasformano un valore di VNC in un punto. Tali tabelle, di cui si riporta un esempio, derivano da una conoscenza del territorio a grande scala (almeno 1:5000).

Id	Via	n.	X	Y
35	Dante Alighieri	1	622100	4724090
36	Dante Alighieri	2	622160	4725001
38	Dante Alighieri	4	622291	4725117
39	Cesare Beccaria	5	622274	4724090
40	Cesare Beccaria	6	622158	4725167
42	Cesare Beccaria	8	622196	4725110

A titolo di esempio riportiamo un frammento di mappa dove è presente uno strato informativo specifico per il geovocabolario; il modello del layer ha, come parte geografica, delle piccole linee e come parte descrittiva quella della tabella precedente.



Il geovocabolario descritto può risultare abbastanza pesante per certe applicazioni, specialmente su dispositivi mobili come un navigatore satellitare; la traduzione da georeferenziazione indiretta a georeferenziazione diretta può avvenire in modo meno accurato ma più semplice considerando, di un tratto di strada, solo i numeri civici iniziale e finale. Il modello dati (semplificato) è descritto di seguito.

Linea -----	Id	NomeVia	Primo sinistra	Ultimo sinistra	Primo destra	Ultimo destra
Linea -----	5	Rosai	13	33	8	20
Linea -----	5	Rosai	35	35	22	26

Si intuisce come, senza bisogno di codificare tutti i civici, la posizione “Via Rosai 15” sia ricavabile per interpolazione. La georeferenziazione è meno accurata, ma comunque molto buona nelle città, dove i tratti di strada da incrocio a incrocio sono abbastanza brevi, l'archivio è più leggero, gestisce senza problemi i civici con esponente e, in molti casi, è meno sensibile in caso di nascita di nuovi civici. L'archivio è addirittura perfetto nelle zone extraurbane, dove a volte la numerazione civica è espressa in metri rispetto ad un'origine arbitraria.