

Linguaggi di interrogazione per basi di dati geografiche

Geo-algebra relazionale

Novembre 2005

Alberto Belussi

L'interrogazione di una base di dati geografica

L'interrogazione di una Base di Dati geografica è influenzata dalla suddivisione dell'informazione territoriale in due componenti: il dato geometrico e il dato alfanumerico. Queste due componenti hanno caratteristiche diverse sia per quanto riguarda la loro memorizzazione sia per le relazioni che esistono nei corrispondenti domini.

In particolare, i dati alfanumerici sono istanze appartenenti a domini monodimensionali, dove è presente una relazione d'ordine. Il dominio dei numeri interi e il dominio delle stringhe di caratteri di lunghezza minore di 10 sono esempi di domini alfanumerici. In essi sono presenti rispettivamente l'usuale relazione d'ordine sugli interi e l'ordinamento alfabetico sulle stringhe.

Sono quindi esprimibili condizioni di selezione che si basano sulla relazione di uguaglianza ($A = B$) o sulla relazione d'ordine ($A < B$, $B \geq A$, ecc.).

L'interrogazione di una base di dati geografica

Le interrogazioni sui dati alfanumerici, memorizzati in una base di dati relazionale, si specificano utilizzando un linguaggio di interrogazione, ad esempio il linguaggio SQL.

I dati geometrici appartengono invece a domini multi-dimensionali (si considerano almeno i valori geometrici del piano cartesiano), dove non esiste una relazione d'ordine di riferimento e sono definiti altri tipi di relazioni di natura spaziale.

La presenza di un comune spazio di riferimento, in cui tutti i valori geometrici sono inseriti, accresce l'importanza delle relazioni di natura spaziale; infatti è attraverso queste ultime che viene espressa la parte geometrica delle interrogazioni.

Relazioni nello spazio

Per poter studiare e classificare le interrogazioni che interessano la parte geometrica dell'informazione territoriale, occorre analizzare le relazioni spaziali che possono esistere tra valori geometrici inseriti nello stesso spazio di riferimento.

Una relazione binaria definisce una proprietà che lega due valori di un determinato insieme. Nel nostro caso i valori sono geometrici e la proprietà riguarda la loro reciproca posizione nello spazio.

Esistono tre classi di relazioni spaziali binarie (vedi dispensa n. 2):

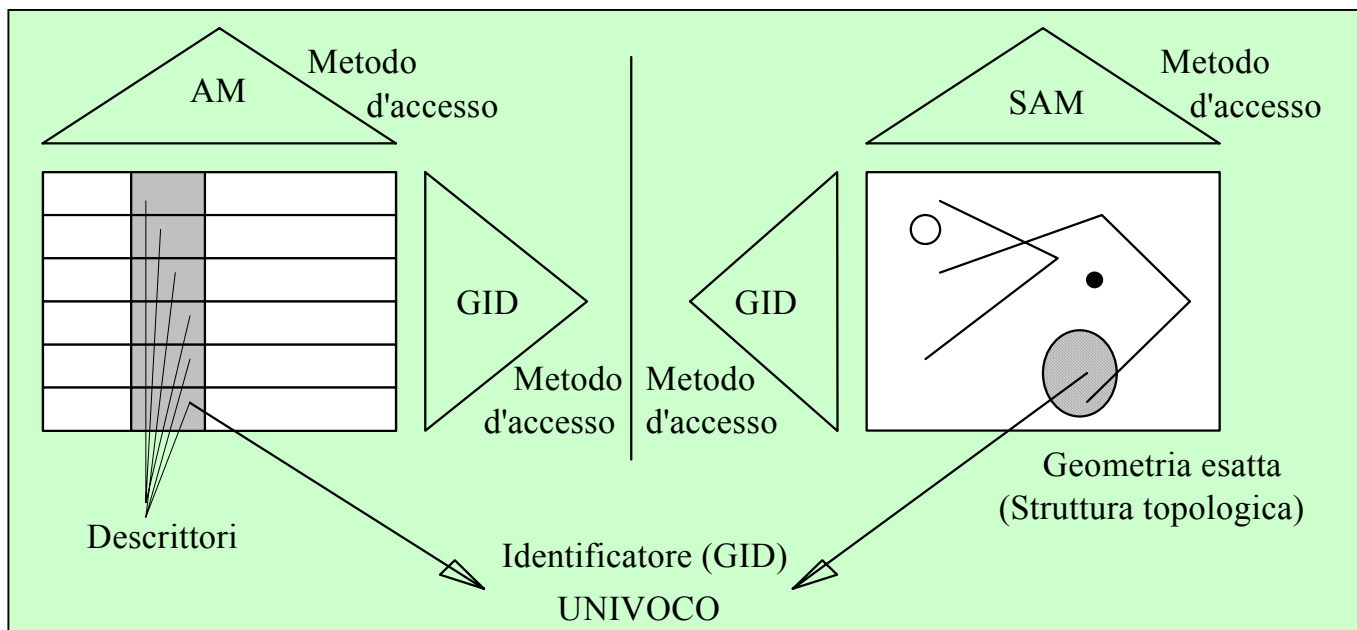
- le **relazioni topologiche**,
- le **relazioni basate sulla direzione** (non realizzabili nel modello “Simple Feature Specification for SQL”),
- le **relazioni basate sulla distanza**.

Interrogazioni in una base di dati geografica

Viste le relazioni spaziali, si presentano ora le categorie di interrogazioni spaziali che ne derivano, dove con interrogazioni spaziali si intendono quelle che selezionano un insieme di valori geometrici in base a operatori di natura spaziale.

Per aver un'idea generale di come viene eseguita un'interrogazione in una base di dati geografica, possiamo fare riferimento allo schema del lucido successivo, dove è mostrata la struttura fisica di una base di dati geografica in un Geo-DBMS.

Interrogazioni in una base di dati geografica



Le interrogazioni puramente geometriche in questa architettura possono produrre come risultato un insieme di valori geometrici “puri” (solo geometria) o un insieme di identificatori per accedere alle informazioni alfanumeriche collegate alla geometria.

Interrogazioni in una base di dati geografica

Esistono essenzialmente due categorie di interrogazioni puramente spaziali:

- selezioni basate su relazioni spaziali rispetto a valori geometrici costanti: questa classe si riferisce alle interrogazioni basate sulle relazioni di ogni valore geometrico con lo spazio di riferimento o con valori geometrici costanti. In particolare in questa categoria vedremo:
 - Range Queries,
 - Selezioni basate sulla distanza.
- join spaziali: questa classe si riferisce alle interrogazioni basate sulle relazioni dei valori geometrici contenuti nella base di dati tra di loro. In particolare in questa categoria vedremo:
 - Join topologici,
 - Join basati sulla distanza.

Range Queries

Le **range query** sono le interrogazioni più diffuse nei sistemi per la gestione di base di dati geografici.

Una range query individua una regione dello spazio di riferimento e seleziona tutti i valori geometrici di uno specifico attributo geometrico della base di dati che sono in una particolare relazione geometrica con la regione specificata.

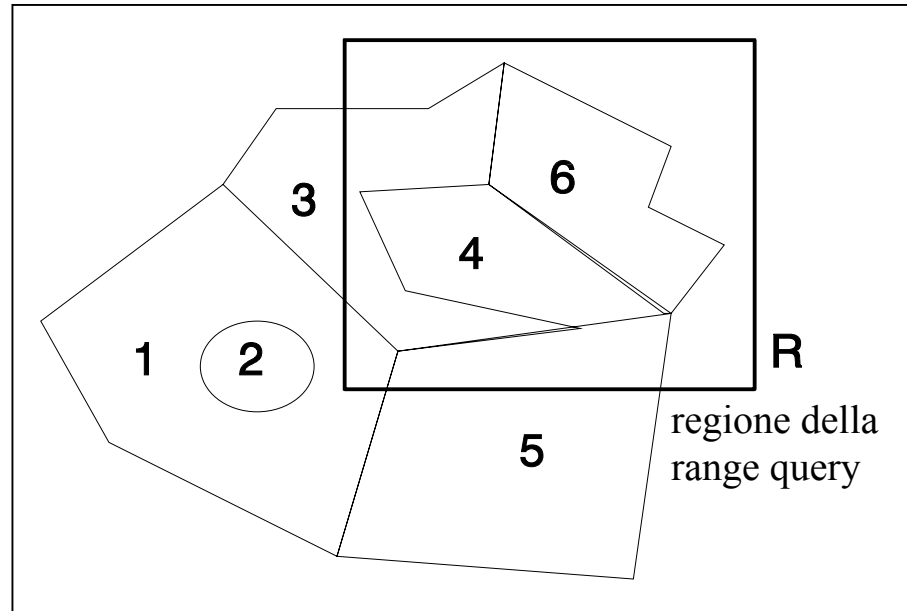
La regione di spazio viene specificata come valore geometrico nell'espressione dell'interrogazione, così come succede nel caso di una costante alfanumerica in una condizione SQL.

La regione di spazio in un range query ha la stessa dimensionalità dello spazio di riferimento. Nel piano cartesiano ad esempio può essere solo un poligono.

La relazione geometrica in base alla quale si selezionano i valori geometrici è topologica e può essere una delle seguenti: “disjoint”, “overlap o cross” e “in”, difficilmente può essere una relazione “touch”.

Range Queries

Esempio



Supponendo di avere i dati in una relazione A con un attributo geometrico g , la range query si potrebbe esprimere come segue:

$$\sigma_{(g.Overlap(R))} (A) \text{ oppure } \sigma_{(g.Within(R))} (A)$$

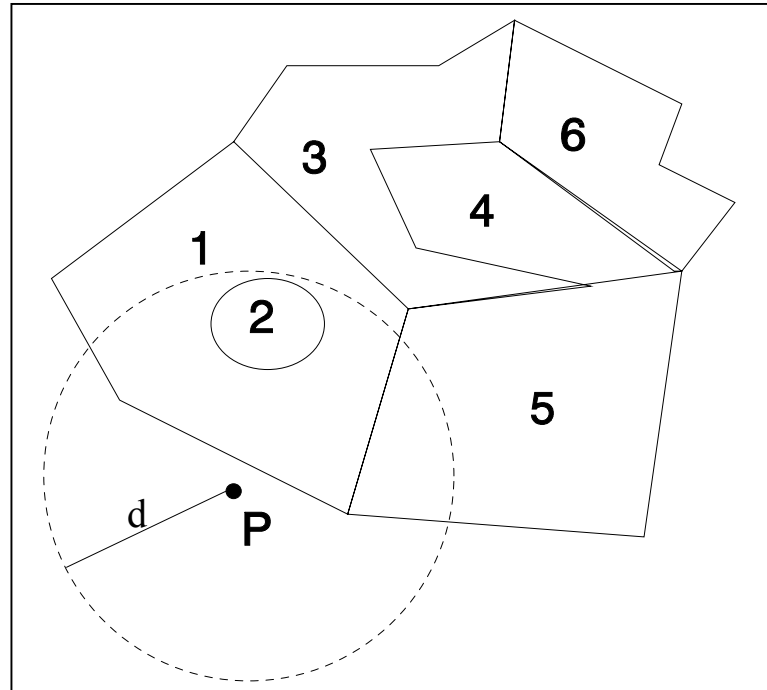
Selezioni basate sulla distanza

Le **selezioni basate sulla distanza** (o buffer query) selezionano tutti i valori geometrici di un attributo geometrico specificato della base di dati che si trovano ad una distanza compresa tra zero e un valore massimo d da un punto o da un valore geometrico specificato.

Questo tipo di query fa quindi riferimento ad una definizione di distanza tra i valori geometrici, che deve essere presente nello spazio di riferimento. Inoltre per l'esecuzione dell'interrogazione deve essere possibile o calcolare la distanza tra due valori geometrici o generare la regione di spazio che contiene tutti i punti dello spazio di riferimento la cui distanza dal punto o valore geometrico di riferimento per la query sia minore di un valore costante d .

Selezioni basate sulla distanza

Esempio



Supponendo di avere i dati in una relazione A con un attributo geometrico g, la range query si potrebbe esprimere come segue:

$$\sigma_{(g.Overlap(P.Buffer(d)))} (A)$$

Estensione dell'operazione di selezione (σ) dell'algebra relazionale

Per poter usare le relazioni presentate in precedenza nell'interrogazione di una base di dati geografica rappresentata nel modello logico di riferimento, è necessario estendere l'operazione di selezione nel seguente modo:

$$\sigma_P(R)$$

dove la condizione di selezione P può essere, oltre che una condizione su attributi alfanumerici che compaiono in R, anche un'espressione integer (1 = TRUE) basata sui metodi disponibili nei tipi geometrici del modello SFS e in particolare sui metodi:

Equals, Disjoint, Touches, Overlaps, Within, Contains, Crosses, Intersects, Relate e Buffer presenti in tutti i tipi geometrici di SFS.

Spatial Joins

Insieme alla range query l'operazione di spatial join è l'operazione spaziale che più frequentemente viene utilizzata nelle interrogazioni di una base di dati geografica.

Lo spatial join è del tutto simile ad un join relazionale dove invece di considerare attributi alfanumerici si considerano attributi geometrici e invece di condizioni alfanumeriche si impongono condizioni basate su relazioni spaziali.

Spatial Joins

A seconda di quali relazioni spaziali vengano utilizzate nella condizione di join, l'operazione prende un nome diverso:

- Join Topologico: in questo caso viene utilizzata una relazione topologica.

Ad esempio, date le tabelle FIUME(idFiume, Nome, Tracciato: Linestring) e COMUNI(idComune, Nome, Estensione: Polygon), attraverso un join topologico si potrebbero identificare tutti i comuni il cui territorio è adiacenti ad un fiume.

L'esecuzione di questo tipo di interrogazioni è velocizzata in presenza di una struttura di memorizzazione degli strati basata sulla topologia.

Spatial Joins

- Join basato sulla direzione: in questo caso si utilizzano le relazioni basate sulla direzione.

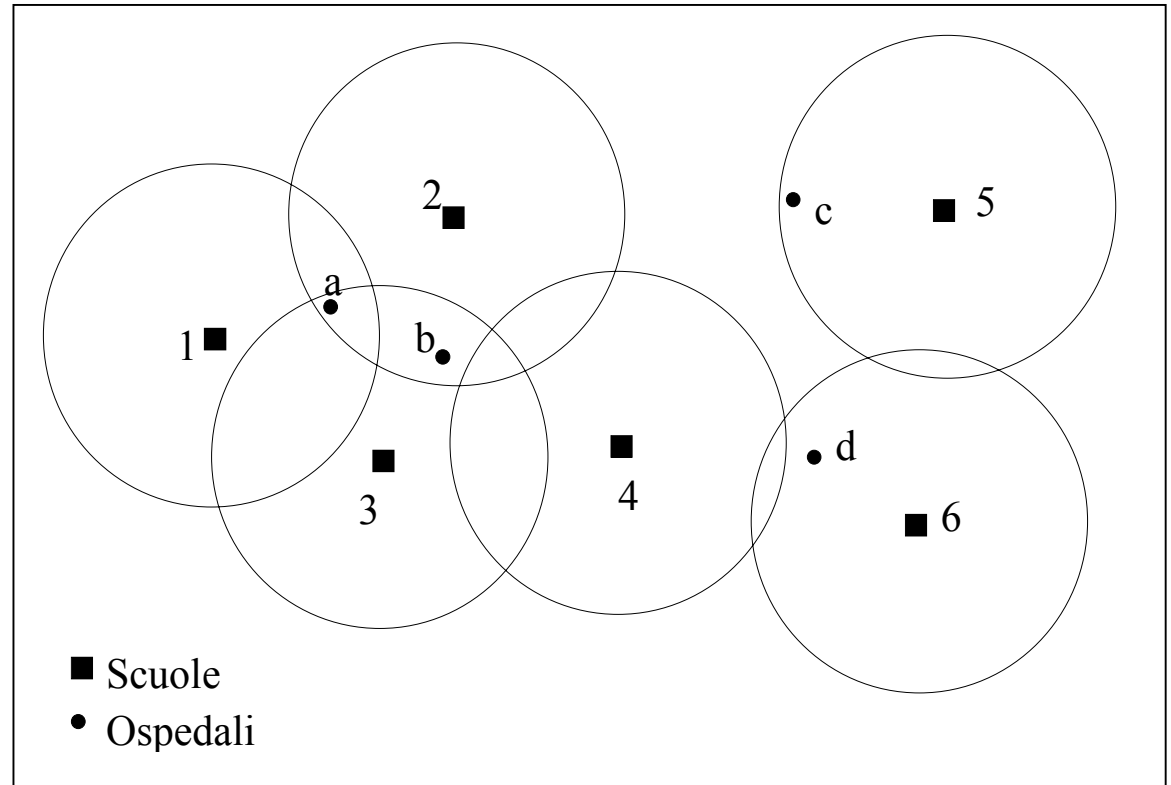
Ad esempio, date le tabelle CIMA(idCima, Nome, Posizione: Point) e RIFUGIO(idRifugio, Nome, Posizione: Point), si potrebbero identificare per ogni cima alpina, tutti i rifugi che non stanno a nord-est.

- Join basato sulla distanza: in quest'ultimo caso si utilizzano le relazioni basate sul concetto di distanza.

Ad esempio, si potrebbero identificare per ogni scuola gli ospedali in un raggio di 2Km.

Spatial Joins

Esempio



Supponendo di avere i dati in due relazioni A (scuole) e B (ospedali) con un attributo geometrico f e g rispettivamente, lo spatial join basato sulla distanza che identifica per ogni scuola gli ospedali nel raggio di 2Km, si potrebbe esprimere come segue:

$$\sigma_{(g.Within(f.Buffer(2Km)))}(A \times B)$$

Applicazione dei metodi associati ai tipi geometrici

Per l'applicazione dei metodi definiti nei tipi geometrici ai valori geometrici contenuti negli attributi geometrici di una relazione si introduce il seguente operatore algebrico:

APPLY DOMAIN FUNCTION (ADF)

$ADF_{a.f(a_1, \dots, a_n), b} (R: \text{relazione}) \rightarrow S: \text{relazione}$

dove:

- $a, a_1, \dots, a_n \in \text{schema}(R)$ e $b \notin \text{schema}(R)$
- $f: D_1 \times \dots \times D_{n+1} \rightarrow D$
- i domini degli attributi a, a_1, \dots, a_n sono D_1, \dots, D_{n+1}

Applicazione dei metodi associati ai tipi geometrici

APPLY DOMAIN FUNCTION (ADF)

La relazione risultato ha le seguenti caratteristiche:

- $\text{schema}(S) = \text{schema}(R) \cup \{b\}$
- il dominio di b è D
- $S = \{s \mid (\exists r)(r \in R \wedge (\forall a')(a' \in \text{schema}(R) \Rightarrow s.a' = r.a') \wedge s.b = r.a.f(r.a_1, \dots, r.a_n))\}$

Esempio

Data la relazione COMUNE(Nome, Estensione: PG)

ADF_{Estensione.Area, AreaC} (COMUNE)