ANNO ACCADEMICO 2003-2004
SISTEMI INFORMATIVI GEOGRAFICI

SISTEMI INFORMATIVI TERRITORIALI (SIT)
GEOGRAPHICAL INFORMATION SYSTEMS (GIS)

Sistemi Informativi Territoriali

5. Un linguaggio di interrogazione per basi di dati geografiche: GEO-ALGEBRA

ALBERTO BELUSSI

NOVEMBRE 2003

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.).

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 multidimensionali (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:

1. LE RELAZIONI TOPOLOGICHE,
2. LE RELAZIONI BASATE SULLA DIREZIONE,
3. LE RELAZIONI BASATE SULLA DISTANZA.

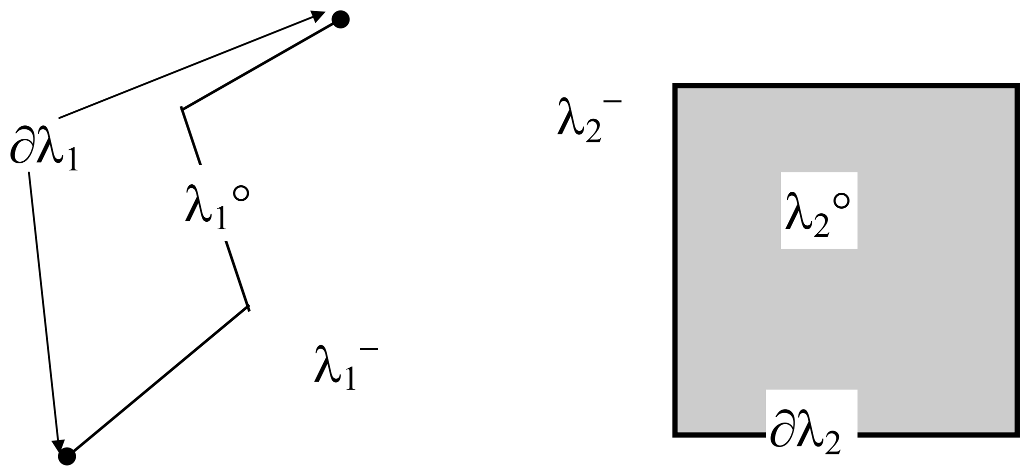
Le relazioni topologiche

Le relazioni topologiche sono definite come quelle relazioni che sono invarianti rispetto a trasformazioni topologiche (rubber sheet transformation). Esse hanno inoltre le seguenti proprietà:

- sono completamente indipendenti dalla distanza tra i due valori geometrici considerati e dall'estensione dei due valori geometrici
- sono di tipo qualitativo: descrivono "come" sono in relazione due valori geometrici senza specificare "quanto" (misura).
- fanno riferimento a concetti di alto livello, che hanno una facile corrispondenza con il linguaggio naturale.

La definizione teorica delle relazioni topologiche fa riferimento a diversi modelli. Fra questi in particolare si è affermato, nell'ambito delle basi di dati geografiche e non solo, il modello di Max Egenhofer (1991).

Tale modello considera come valori geometrici insiemi di punti chiusi e si basa sulla suddivisione dello spazio prodotta da ogni valore geometrico λ. Tale suddivisione individua: l'**interior** λ° (parte interna), il **boundary** $\partial\lambda$ (frontiera) e l'**exterior** λ^- (esterno o complemento) di un valore geometrico λ.
Ad esempio:



In tale modello tutte le relazioni topologiche possibili tra due valori geometrici A e B si definiscono attraverso il risultato dell'intersezione tra tutte le suddivisioni dello spazio generate da A (A°, ∂A, A^-) e tutte le suddivisioni dello spazio generate da B (B°, ∂B, B^-). Formalmente:

$$R(A,B) = \begin{pmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap B^- \\ \partial A \cap B^\circ & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^\circ & A^- \cap \partial B & A^- \cap B^- \end{pmatrix}$$

Il numero totale di relazioni è pari a 512, anche se non tutte le combinazioni corrispondono a casi reali.

Il modello è applicabile a:

- due linee nel piano cartesiano
- due poligoni nel piano cartesiano
- una linea e un poligono nel piano cartesiano
- due poligoni nello spazio a tre dimensioni
- una linea e un poliedro nello spazio a tre dimensioni
- due poliedri nello spazio a tre dimensioni

Se consideriamo il caso di due poligoni nel piano cartesiano basta considerare quattro intersezioni per avere tutti i casi significativi:

$$\begin{matrix} A^\circ \cap B^\circ & A^\circ \cap \partial B \\ \partial A \cap B^\circ & \partial A \cap \partial B \end{matrix}$$

Le relazioni topologiche

E. Clementini, P. Di Felice e P. van Oosterom hanno proposto un insieme di relazioni topologiche applicabili a tutte le tipologie (punto, linea, poligono) di valori geometrici.

Dati due valori geometrici λ_1 e λ_2 si definiscono le seguenti relazioni topologiche:

TOUCH

$$\lambda_1 \text{ TOUCH } \lambda_2 \quad \Leftrightarrow \quad (\lambda_1° \cap \lambda_2° = \emptyset) \wedge (\lambda_1 \cap \lambda_2 \neq \emptyset)$$

IN

$$\lambda_1 \text{ IN } \lambda_2 \quad \Leftrightarrow \quad (\lambda_1° \cap \lambda_2° \neq \emptyset) \wedge (\lambda_1 \cap \lambda_2 = \lambda_1) \wedge (\lambda_1 \cap \lambda_2 \neq \lambda_2)$$

<: cross tra poligono e un insiemi finito di punti.

CROSS

$$\lambda_1 \text{ CROSS } \lambda_2 \quad \Leftrightarrow \quad dim(\lambda_1° \cap \lambda_2°) \leq (max(dim(\lambda_1), dim(\lambda_2)) - 1) \wedge (\lambda_1 \cap \lambda_2 \neq \lambda_1) \wedge (\lambda_1 \cap \lambda_2 \neq \lambda_2)$$

OVERLAP

$$\lambda_1 \text{ OVERLAP } \lambda_2 \Leftrightarrow \quad dim(\lambda_1° \cap \lambda_2°) = dim(\lambda_1) = dim(\lambda_2) \wedge (\lambda_1 \cap \lambda_2 \neq \lambda_1) \wedge (\lambda_1 \cap \lambda_2 \neq \lambda_2)$$

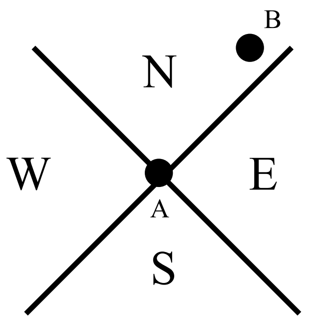
DISJOINT

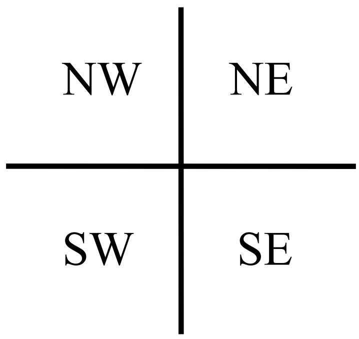
$$\lambda_1 \text{ DISJOINT } \lambda_2 \Leftrightarrow \quad (\lambda_1 \cap \lambda_2 = \emptyset)$$

dove la funzione *dim* restituisce la dimensione del risultato: 0 punto, 1 linea, 2 poligono oppure *empty* se l'intersezione è vuota.

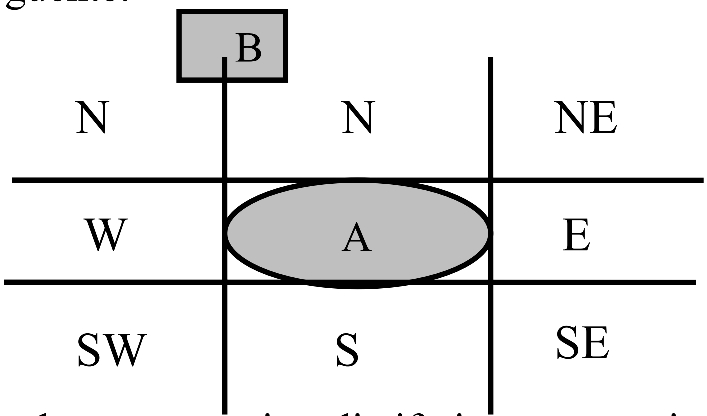
Le relazioni basate sulla direzione

Le relazioni direzionali sono definite sulla base di un sistema di riferimento direzionale che partiziona lo spazio almeno in quattro quadranti. L'origine del sistema di riferimento viene posta su uno dei due valori geometrici considerati. Intersecando i quadranti con l'altro valore geometrico si deriva la relazione tra i due valori.
Considerando valori geometrici puntiformi, le partizioni dello spazio possono essere ad esempio le seguenti:





Considerando invece valori geometrici con estensione significativa il modello da adottare deve tener conto di tale estensione e quindi può essere ad esempio il seguente:



Preso A come valore geometrico di riferimento, se si applica il modello topologico per descrivere le relazioni tra le partizioni dello spazio e l'altro valore geometrico B, si ottengono 169 relazioni.

Le relazioni basate sulla distanza

Queste relazioni richiedono la definizione nello spazio di riferimento di una funzione che calcola la distanza. Se si utilizza la distanza Euclidea allora la distanza tra due punti $P=(x_1,y_1)$ $Q=(x_2,y_2)$ è definita come segue:

$$D(P, Q) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

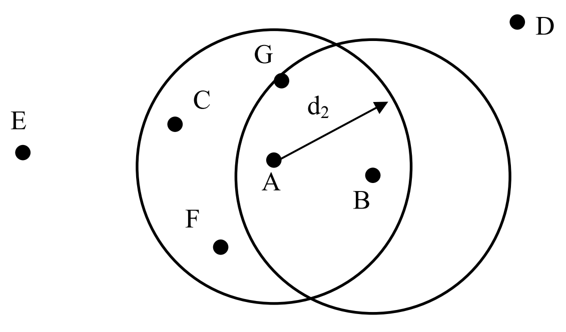
Se consideriamo invece due valori geometrici con estensione significativa, occorre ridefinire il concetto di distanza, ad esempio come segue:

$$\mathit{Dist}(A, B) = \mathit{min}(\{D(P,Q): P \in A, Q \in B\})$$

Fissata quindi una coppia di distanze (d_1,d_2) è possibile definire una relazione metrica tra due valori geometrici come segue:

$$R_{(d_1,d_2)}(A, B) \Leftrightarrow (d_1 < \mathit{Dist}(A,B) < d_2)$$

Ad esempio, considerando la relazione $R_{(0,d_2)}$:



Sono in relazione con A i punti: B, C, G, F.

Relazioni approssimate

Può essere necessario rappresentare relazioni tra valori geometrici che non derivano da misure precise di distanza, ma che si basano sui concetti astratti di vicinanza e lontananza usualmente presenti nel linguaggio naturale.

Tali relazioni sono dette **relazioni metriche approssimate**. Un insieme di relazioni metriche approssimate ha cardinalità finita, ad esempio: {same_location, very_near, near, medium, far, very_far} costituisce un insieme di relazione metriche approssimate.

E' solitamente possibile fissare un ordinamento delle relazioni, ad esempio:

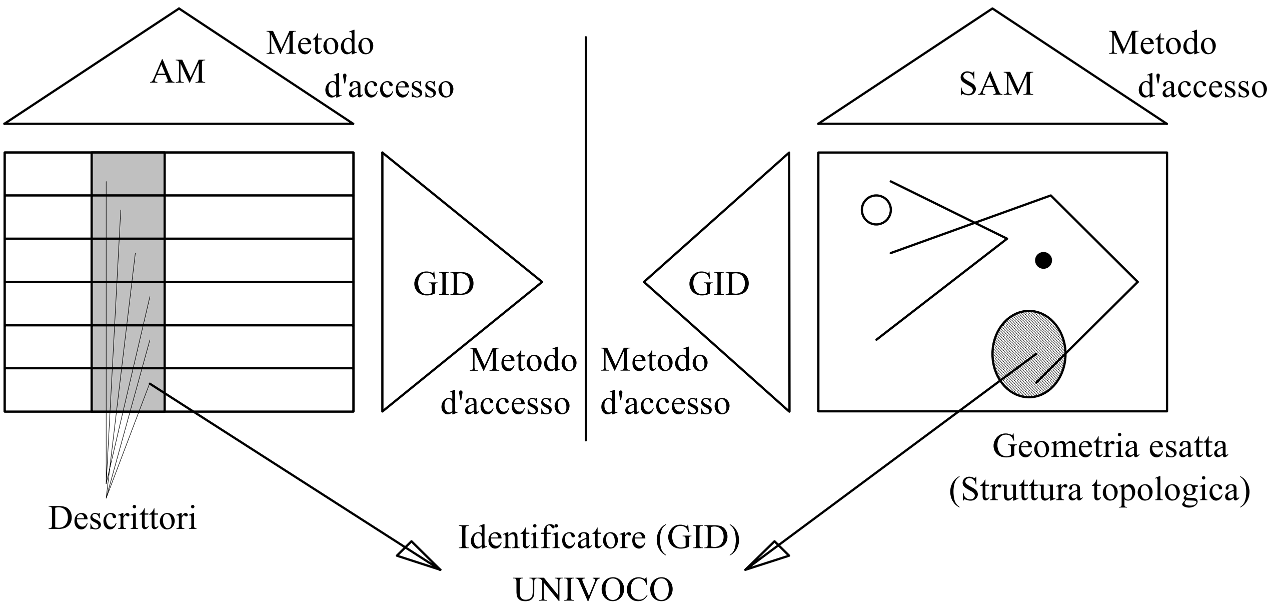
same_location < very_near < near < medium < far < very_far

Quindi un insieme di relazioni metriche approssimate partiziona completamente lo spazio di riferimento. La definizione precisa di tale partizione dipende dal contesto (c'è differenza tra il "near" di chi cammina e il "near" di chi viaggia in aereo) e comunque può essere precisata attraverso degli intervalli di distanza.

Le 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 alla seguente figura, dove è mostrata la struttura di un GEO-DBMS (architettura ibrida).



Le interrogazioni puramente geometriche in questa struttura 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.

Esistono essenzialmente due categorie di interrogazioni puramente spaziali:

1. SELEZIONI BASATE SU RELAZIONI SPAZIALI RISPETTO A VALORI GEOMETRICI COSTANTI: questa classe si riferisce alle interrogazioni basate sulla 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 DIREZIONE,
    - SELEZIONI BASATE SULLA DISTANZA.

2. 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 DIREZIONE,
    - JOIN BASATI SULLA DISTANZA.

Range Query

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.

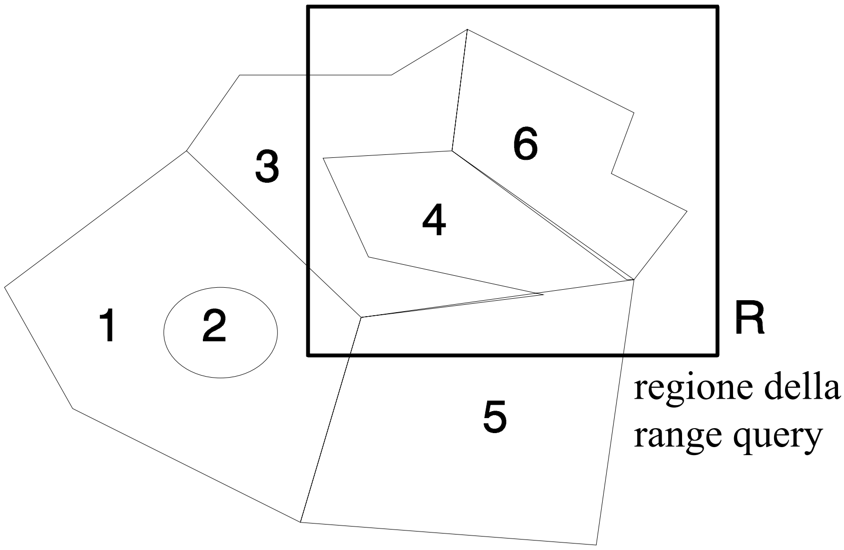
La regione di spazio in un range query ha la stessa dimensionalità dello spazio di riferimento. Nel piano cartesiano ad esempio può essere solo un poligono.

La relazione geometrica in base alla quale si selezionano i valori geometrici è topologica e può essere una delle seguenti:

- la relazione "disjoint",
- la relazione "overlap o cross",
- la relazione "in".

difficilmente può essere una relazione "touch".

Ad esempio:



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

$$\sigma_{(\text{R OVERLAP A.g})}\ (\text{A})$$

oppure

$$\sigma_{(\text{A.g IN R})}\ (\text{A})$$

Selezioni basate sulla direzione

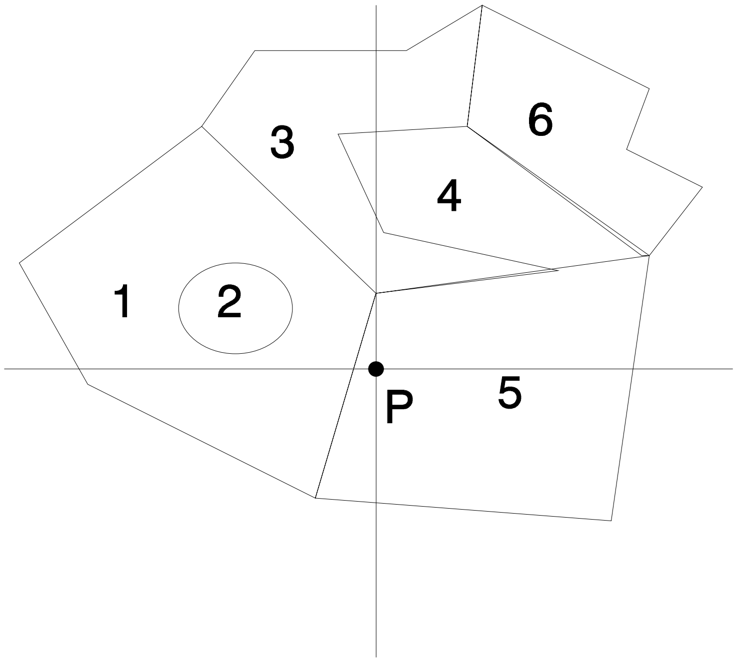
La selezione in base a **proprietà di direzione** dipende dalla specifica di:

- un punto o valore geometrico con estensione,
- un modello delle direzioni,
- una relazione topologica.

L'interrogazione restituisce tutti i valori geometrici contenuti in un attributo geometrico nella base di dati che sono in una particolare relazione topologica con una delle suddivisioni dello spazio derivanti dal riferimento direzionale (rosa dei punti cardinali) posizionato sul punto o sul valore geometrico specificato.

E' possibile anche combinare con i connettivi logici diverse condizioni di questo tipo.

Ad esempio:



Supponendo di avere i dati in una relazione A con un attributo geometrico g, una interrogazione basata sulla direzione potrebbe essere la seguente:

$$\sigma_{(A.g\ IN\ P.NE)}\ (A)$$

oppure

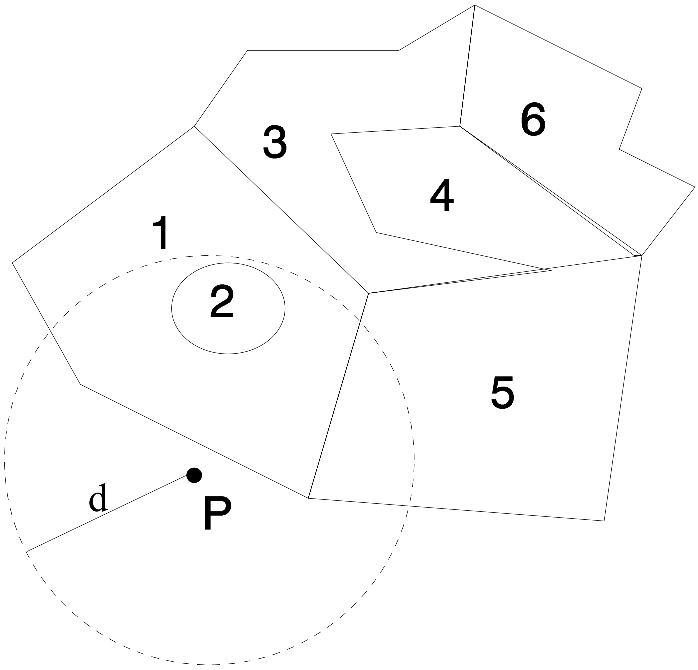
$$\sigma_{(P.NE\ OVERLAP\ A.g)}\ (A)$$

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 un valore minimo d_1 e un valore massimo d_2 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 rientri nei limiti definiti da d_1 e d_2.

Ad esempio,



Supponendo di avere i dati in una relazione A con un attributo geometrico g, una selezione basata sulla distanza potrebbe essere la seguente:

$$\sigma_{\text{(Buffer(P, 0, d) OVERLAP A.g)}}\ (A)$$

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 una condizione basata su:

- <u>relazioni topologiche</u>: $(g\ \theta f)$ oppure $(g\ \theta A)$: dove g e f sono attributi geometrici di R, $\theta \in \{TOUCH, IN, CONTAINS, EQUAL, CROSS, OVERLAP, DISJOINT\}$ e A è un valore geometrico costante.

- <u>relazioni topologiche e funzione buffer</u>: supponiamo che sia presente una funzione $buffer(x, d_1, d_2)$ la quale, dato un valore geometrico qualsiasi x, restituisce un valore geometrico di tipo POLYGON che rappresenta l'insieme dei punti che si trovano ad una distanza d da x dove:

$$d_1 \le d \le d_2$$

Esempi:

$$\sigma_{g\ TOUCH\ f}(R)$$

$$\sigma_{g\ OVERLAP\ Buffer(f,0,10Km)}(R)$$

$$\sigma_{f\ IN\ S}(R)$$

dove g e f sono attributi geometrici di R e S è un valore geometrico costante di tipo POLYGON.