Corso

di

Basi di Dati Spaziali

Linguaggi:
(Geo) Algebra relazionale

Angelo Montanari

Donatella Gubiani

Linguaggi per basi di dati

- Si possono distinguere diversi tipi di linguaggi per la gestione dei dati:
  - Linguaggio di definizione dei dati (DDL)
  - Linguaggio di manipolazione dei dati (DML)
    - Linguaggio di aggiornamento
    - Linguaggio di interrogazione

ALGEBRA RELAZIONALE

- È un linguaggio di interrogazione dei dati di tipo procedurale
- Comprende un insieme di operatori
  - su relazioni
  - che producono relazioni
  - e possono essere composti

Operazioni

- Unarie
  - selezione
  - proiezione
  - rinomina
- Binarie (o insiemistiche)
  - unione
  - differenza insiemistica
  - prodotto cartesiano
- Derivate
  - intersezione
  - join (join naturale, theta-join)

Per trattare dati spaziali…

Dato un linguaggio di interrogazione (algebra relazionale)

→ vengono introdotti nuovi costrutti per il trattamento di dati spaziali (GEO-algebra)

Relazioni Spaziali - 1

- Per poter studiare e classificare le interrogazioni spaziali occorre analizzare le relazioni spaziali
- Una relazione binaria definisce una proprietà che lega due valori di un determinato insieme
  - i valori sono geometrici
  - la proprietà riguarda la loro posizione relativa nello spazio

Relazioni Spaziali - 2

- Tre classi di relazioni spaziali binarie:
  - relazioni topologiche
  - relazioni basate sulla direzione
  - relazioni basate sulla distanza

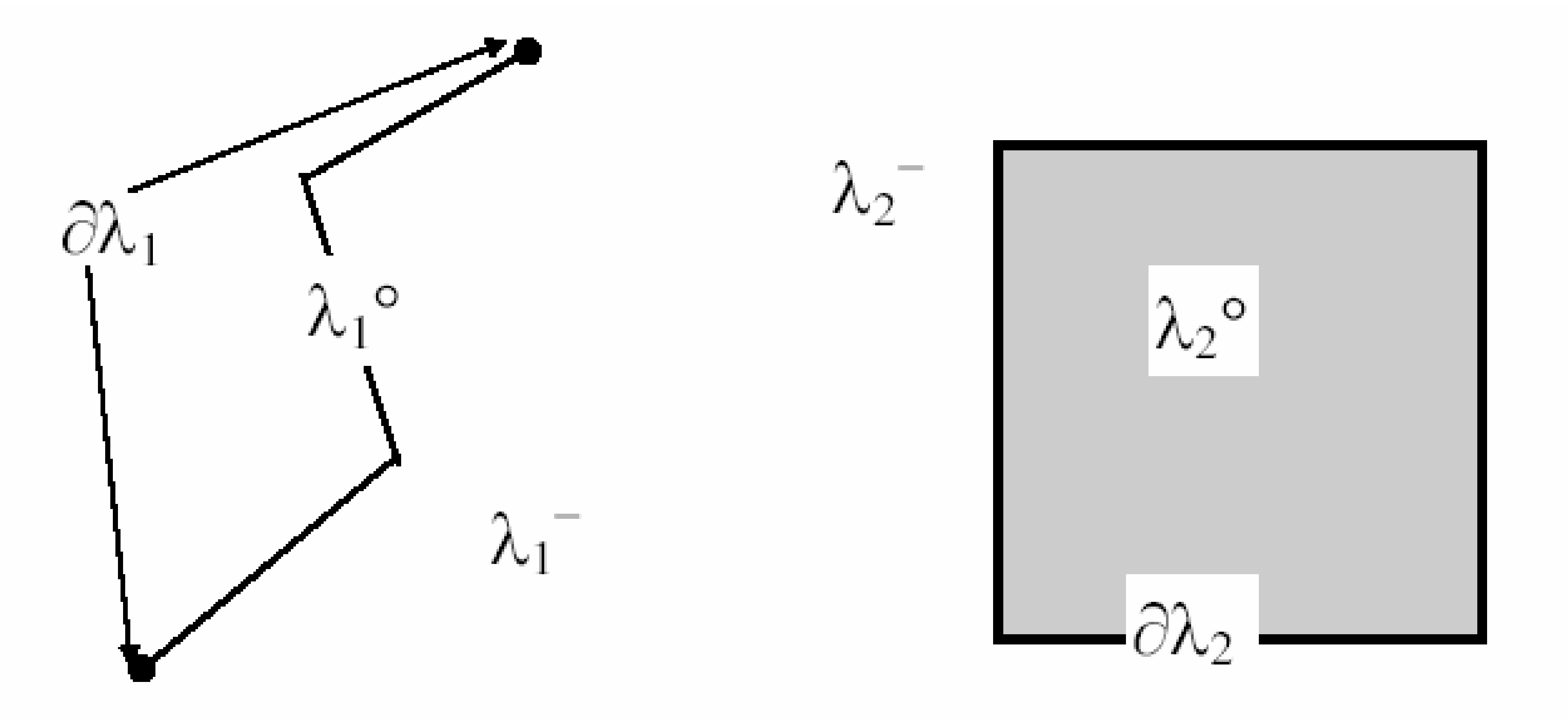
Relazioni topologiche - 1

- Esplicitano il legame topologico esistente fra due entità spaziali
  - esempio: il Tevere attraversa Roma
- Proprietà:
  - indipendenti dalla distanza e dall’estensione
  - di tipo qualitativo
  - fanno riferimento a concetti di alto livello

Relazioni topologiche - 2

- Modello di M. Egenhofer (1991)
  - considera come valori geometrici insiemi di punti chiusi
  - si basa sulla suddivisione dello spazio prodotta da ogni valore geometrico che individua:
    - Interior ($\lambda°$): parte interna
    - Boundary ($\delta\lambda$): frontiera
    - Exterior (λ^-): parte esterna

Relazioni topologiche - 3



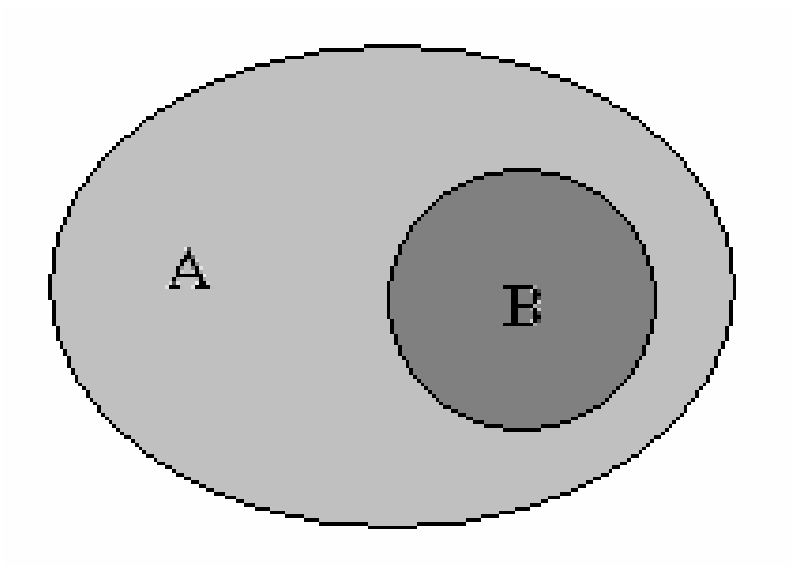
Relazioni topologiche - 4

- Tutte le relazioni topologiche possibili fra due valori geometrici A e B sono definite attraverso il risultato dell'intersezione tra tutte le suddivisioni dello spazio generate da A e tutte le suddivisioni dello spazio generate da B

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

Relazioni topologiche - 5

- Esempio: A contiene B



$$R(A,B)=\begin{pmatrix} \neg\varnothing & \neg\varnothing & \neg\varnothing \\ \varnothing & \varnothing & \neg\varnothing \\ \varnothing & \varnothing & \neg\varnothing \end{pmatrix}$$

Relazioni topologiche - 6

- Il numero totale di relazioni rappresentabili è 512 anche se non tutte sono significative
- 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 tridimensionale
  - una linea e un poliedro nello spazio tridimensionale
  - due poliedri nello spazio tridimensionale

Relazioni topologiche - 7

- E. Clementi, 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

Relazioni topologiche - 8

- Dati due valori geometrici λ_1 e λ_2 di definiscono le seguenti relazioni topologiche:
  - λ_1 **TOUCH** $\lambda_2 \Leftrightarrow (\lambda_1°\cap\lambda_2° =\emptyset)\wedge(\lambda_1\cap\lambda_2\neq\emptyset)$
  - λ_1 **IN** $\lambda_2 \Leftrightarrow (\lambda_1°\cap\lambda_2° \neq\emptyset) \wedge (\lambda_1\cap\lambda_2=\lambda_1) \wedge$
    $\wedge (\lambda_1\cap\lambda_2\neq\lambda_2)$

Relazioni topologiche - 9

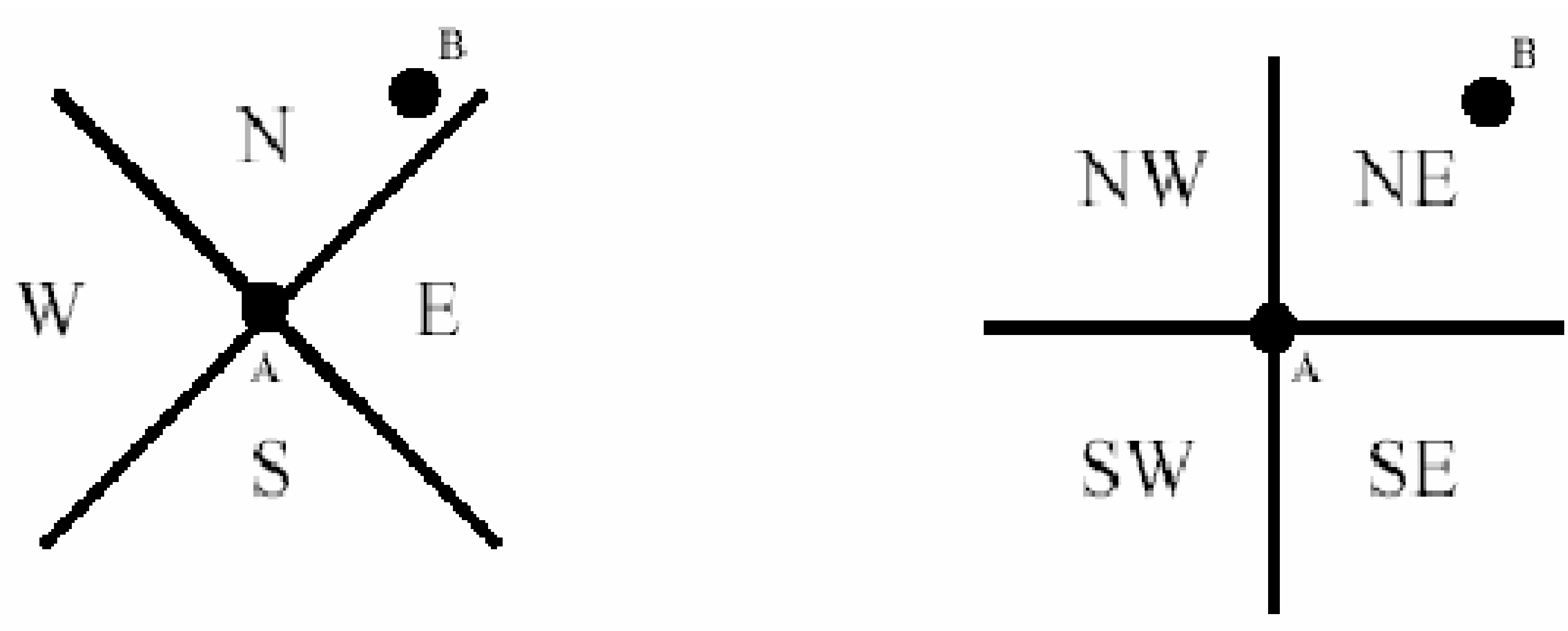
- λ_1 **CROSS** $\lambda_2 \Leftrightarrow$ $\mathit{dim}(\lambda_1° \cap \lambda_2°) \leq (\mathit{max}(\mathit{dim}(\lambda_1), \mathit{dim}(\lambda_2)) - 1) \wedge$ $\wedge (\lambda_1 \cap \lambda_2 \neq \lambda_1) \wedge (\lambda_1 \cap \lambda_2 \neq \lambda_2)$
- λ_1 **OVERLAP** $\lambda_2 \Leftrightarrow$ $\mathit{dim}(\lambda_1° \cap \lambda_2° \neq \emptyset) = \mathit{dim}(\lambda_1) = \mathit{dim}(\lambda_2) \wedge$ $\wedge (\lambda_1 \cap \lambda_2 \neq \lambda_1) \wedge (\lambda_1 \cap \lambda_2 \neq \lambda_2)$
- λ_1 **DISJOINT** $\lambda_2 \Leftrightarrow (\lambda_1 \cap \lambda_2 = \emptyset)$

Note: *dim* restituisce la dimensione del risultato (0 punto, 1 linea, 2 poligono, o empty se l'intersezione è vuota)

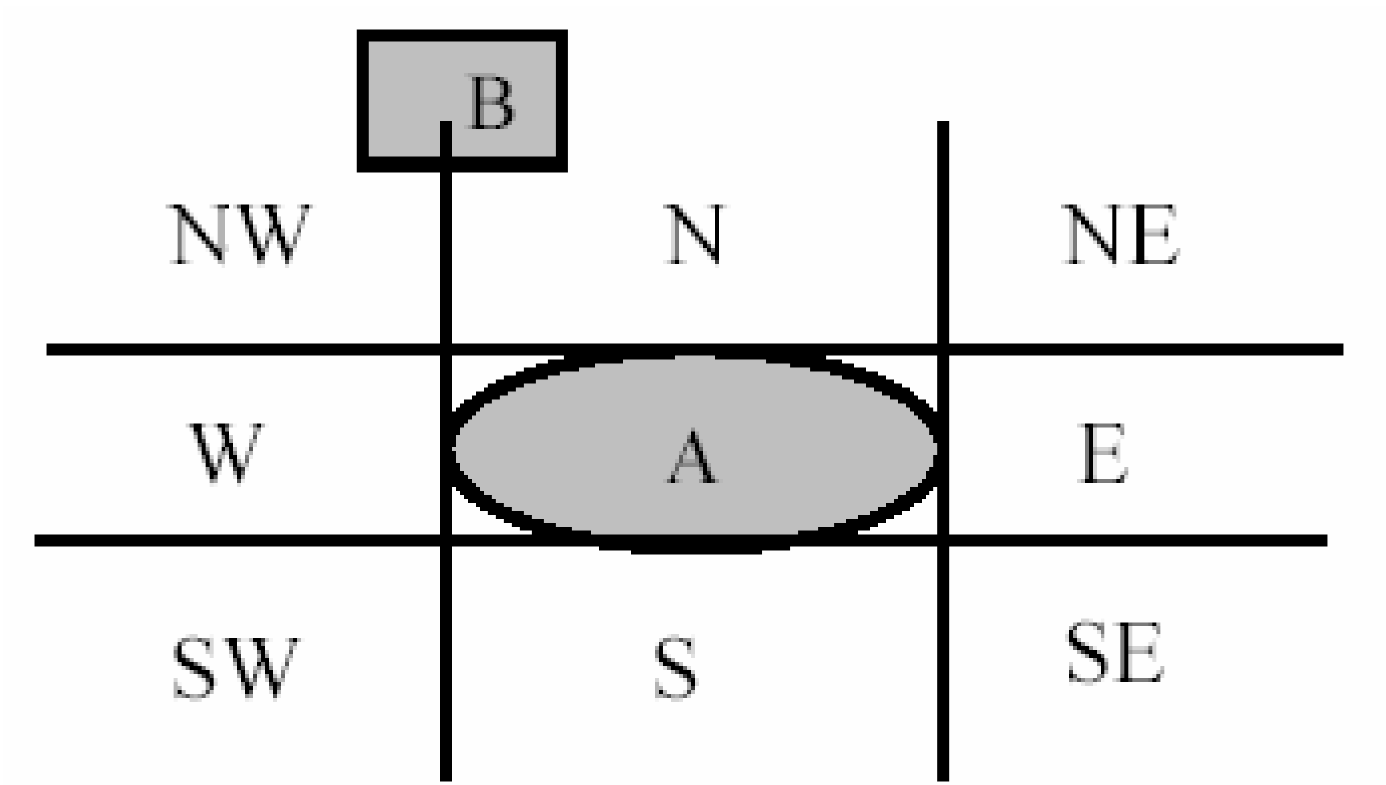
Relazioni basate sulla direzione - 1

- Specificano la posizione in cui un oggetto si trova rispetto ad un'altro
  - esempio: Udine è a nord-est di Venezia
- Sono definite sulla base di un sistema di riferimento direzionale che partiziona lo spazio in almeno quattro quadranti
  - l'origine del sistema viene posto su uno dei due valori geometrici considerati
  - intersecando i quadranti con l'altro valore geometrico si deriva la relazione tra i due valori

Relazioni basate sulla direzione - 2



Relazioni basate sulla direzione - 3



Relazioni basate sulla distanza - 1

- Specificano la distanza fra due oggetti
  - esempio: la piscina è a 1 km dall'ospedale
- È necessaria la definizione nello spazio di riferimento di una funzione che calcola la distanza
  - distanza euclidea fra due punti $P=(x_1,y_1)$ e $Q=(x_2,y_2)$

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

Relazioni basate sulla distanza - 2

– se si considerano due valori con estensione significativa occorre ridefinire il concetto di distanza, ad esempio:

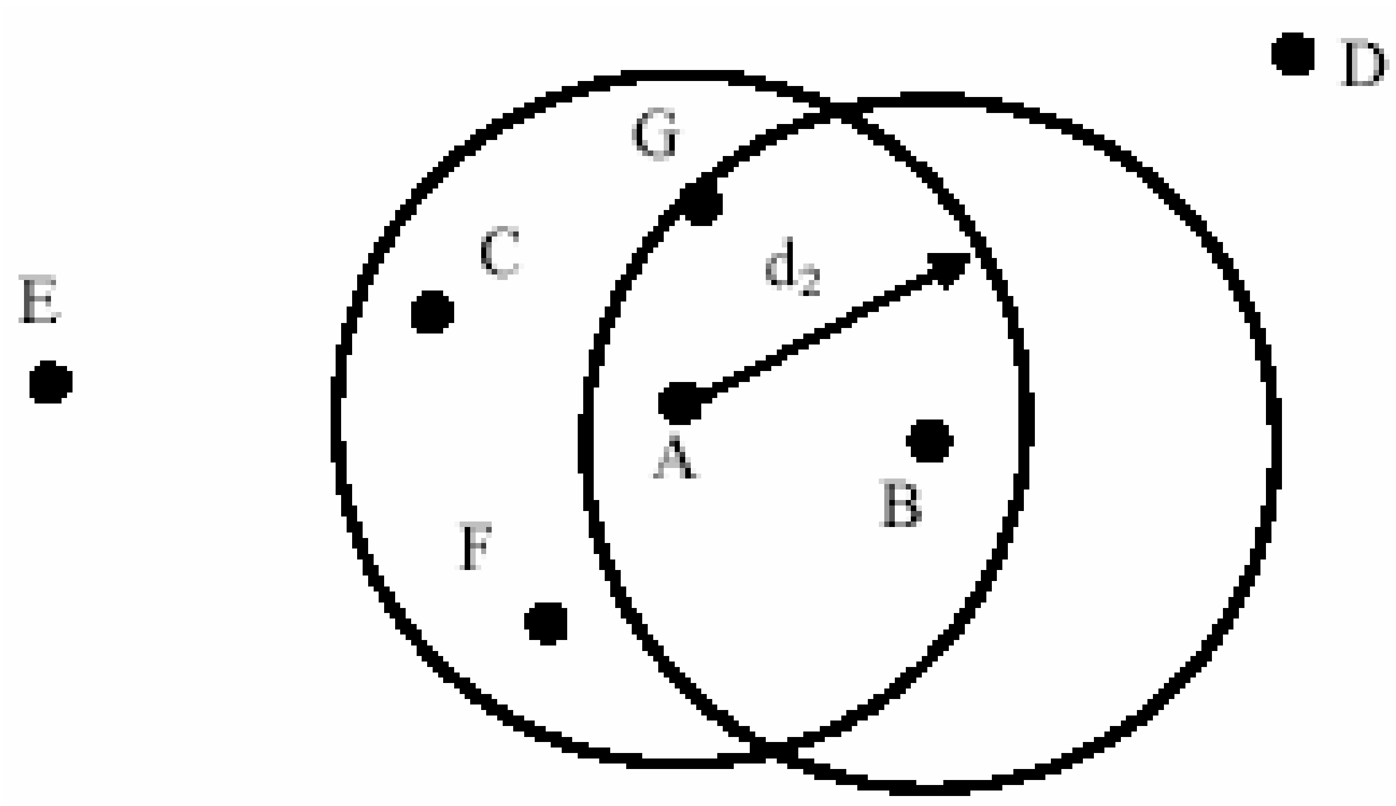
$$\mathit{Dist}(A,B)=\mathit{min}(\{D(P,Q):P\in A,Q \in B\})$$

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

$$R_{(d1,d2)}(A,B) \Leftrightarrow (d_1<\mathit{Dist}(A,B)<d_2)$$

Relazioni basate sulla distanza - 3

- Esempio: $R(0,d_2)$



Interrogazione di una base di dati spaziale

- Estensione dell'algebra relazionale che introduce costrutti per interrogazioni spaziali
- Esistono due categorie di interrogazioni puramente spaziali:
  - selezioni basate su relazioni spaziali rispetto a valori geometrici costanti
  - join spaziali

Selezioni basate su relazioni spaziali rispetto a valori geometrici costanti

- Si riferiscono alle interrogazioni basate sulle relazioni di ogni valore geometrico con lo spazio di riferimento o con valori geometrici costanti

- In questa categoria vedremo:
  - range query
  - selezioni basate sulla direzione
  - selezioni basate sulla distanza

Sulle condizioni di selezione

Dato che ogni tabella (tema) possiede un solo attributo geometrico, l'unica possibile tipologia di condizione di selezione confronta il valore di tale attributo con un valore geometrico costante (o con lo spazio di riferimento)

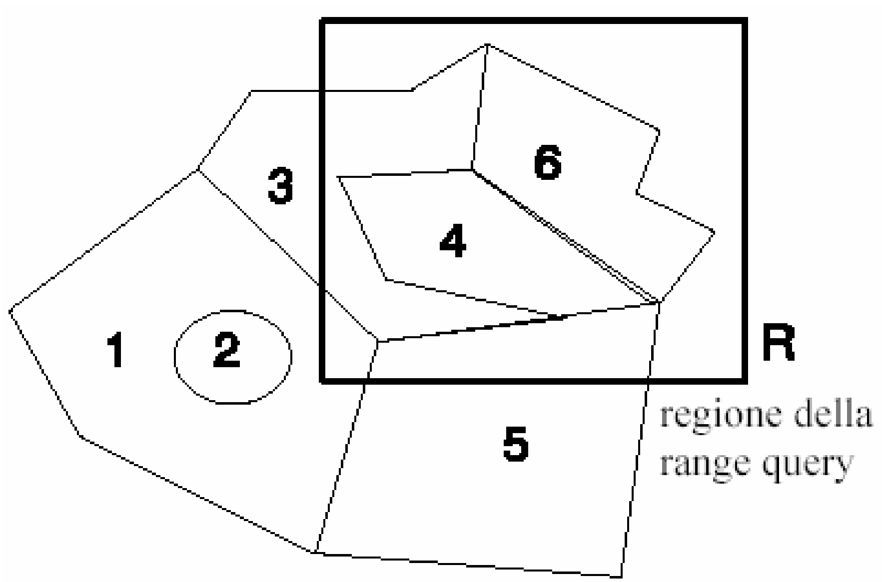
Per confrontare i valori geometrici di tuple diverse di una stessa tabella, occorre duplicare la tabella e utilizzare l'operazione di join spaziale

Range query - 1

- Seleziona tutti i valori geometrici della base di dati che sono in una particolare relazione con una regione dello spazio di riferimento
- Le relazioni topologiche di riferimento sono:
  - disjoint
  - overlap
  - cross
  - in/contain
  - touch
  - equal

Range query - 2

- Dati una tabella A con attributo geometrico g e una regione R



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

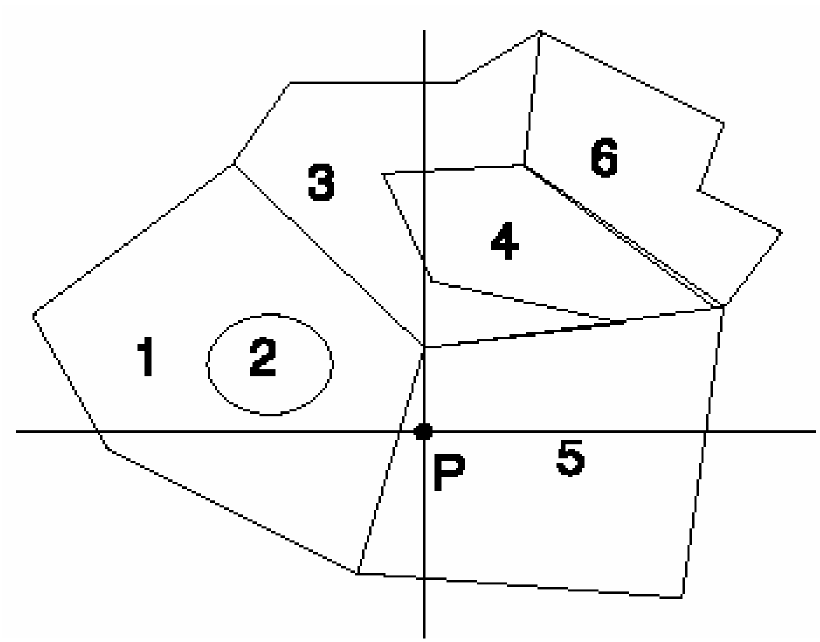
$$\sigma_{(A.g \ \ IN \ \ R)}(A)$$

Selezioni basate sulla direzione - 1

- Seleziona tutti i valori geometrici della base di dati che sono in una particolare relazione topologica con una delle suddivisioni dello spazio derivanti dal riferimento direzionale posizionato su un particolare valore geografico

Selezioni basate sulla direzione -2

- Dati una tabella A con attributo geometrico g e un valore geometrico P (punto)



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

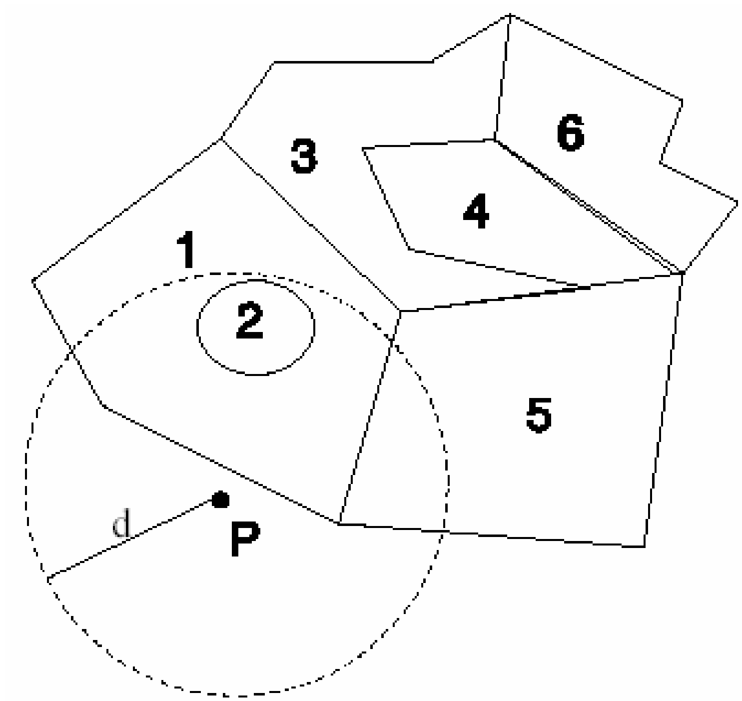
$$\sigma_{(A.g \quad IN \quad P.NE)}(A)$$

Selezioni basate sulla distanza - 1

- Seleziona tutti i valori geometrici della base di dati che sono ad una distanza compresa tra un valore minimo d_1 e un valore massimo d_2 da un particolare valore geografico

Selezioni basate sulla distanza - 2

- Dati una relazione A con attributo geometrico g, un valore geometrico P (punto) e due distanze 0 e d



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

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

Join spaziali

- Si riferiscono alle interrogazioni basate sulle relazioni dei valori geometrici contenuti nella base di dati
- In questa categoria vedremo:
  - join topologici
  - join basati sulla direzione
  - join basati sulla distanza

Join topologici

- Date le tabelle FIUME(nome,tracciato:LN) e COMUNE(nome,estensione:PG) identificare tutti i comuni il cui territorio è attraversata da un fiume

$$FIUME \underset{tracciato \ \ CROSS \ \ estensione}{\bowtie} COMUNE$$

Join basati sulla direzione

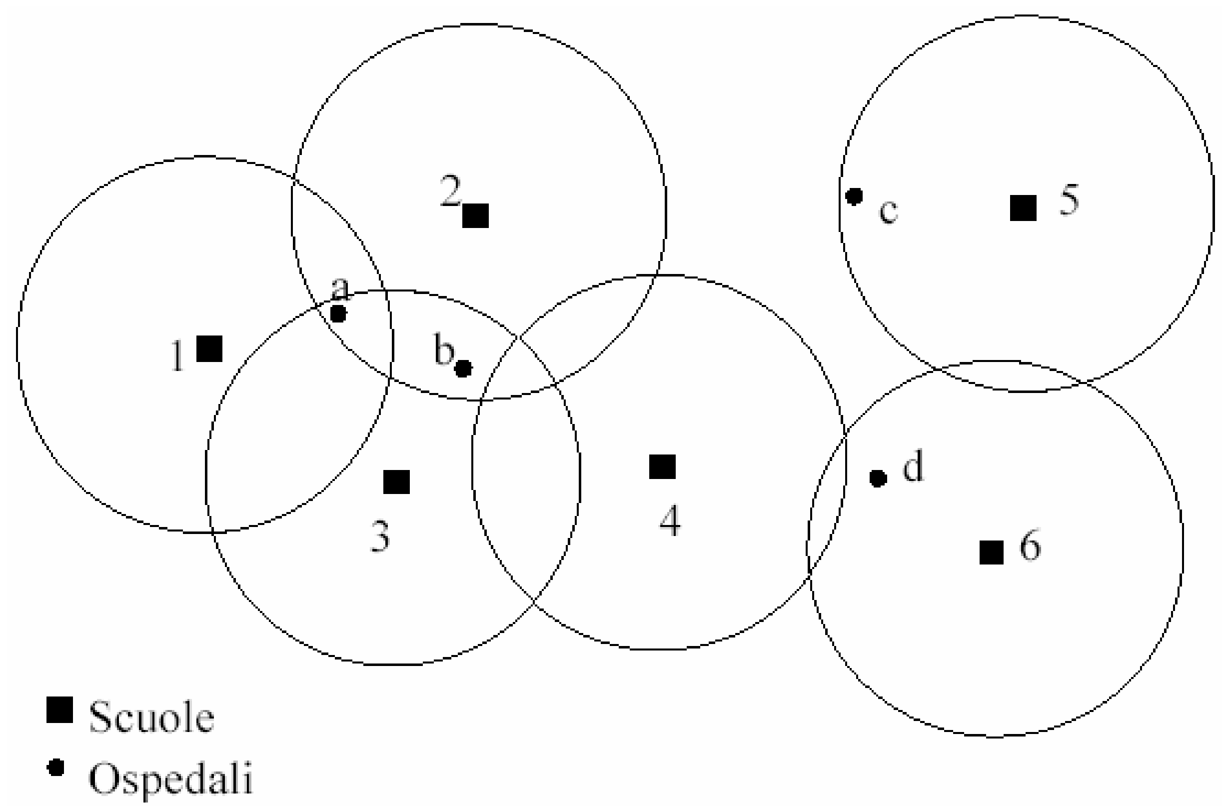
- Date le tabelle CIMA(nome,posizioneC:PT) e RIFUGIO(nome,posizioneR:PT) identificare per ogni cima tutti i rifugi che stanno a nord-est

$$RIFUGIO \underset{posizioneR \ \ IN \ \ posizioneC.NE}{\bowtie} CIMA$$

Join basati sulla distanza - 1

- Date le tabelle OSPEDALE(nome,posizioneO:PT) e SCUOLA(nome,posizioneS:PT) identificare per ogni scuola gli ospedali in un raggio di 2Km

Join basati sulla distanza - 2



$$OSPEDALE \underset{posizioneO \quad IN \quad Buffer(posizioneS, 0{,}2Km)}{\bowtie} SCUOLA$$

In conclusione…

- Per poter usare le relazioni spaziali è necessario estendere le operazioni di selezione e di join in modo che le condizioni possano includere anche:
  - relazioni topologiche: ($g\theta f$) o ($g\theta A$), dove g ed f sono attributi geometrici, $\theta\in$ {TOUCH, IN, CONTAINS, EQUAL, CROSS, OVERLAP, DISJOINT} e A è un valore geometrico costante
  - relazioni topologiche e funzione buffer: dove *buffer*(x,d_1,d_2), applicata ad 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 $d_1<d<d_2$

Applicazione dei metodi ai tipi geometrici - 1

- Si introduce l'operatore algebrico APPLY DOMAIN FUNCTION (ADF)

$$ADF_{a.f(a1\ldots an),b}(R{:}relazione) \rightarrow S{:}relazione$$

dove

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

La relazione risultato ha le seguenti caratteristiche:

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

Applicazione dei metodi ai tipi geometrici - 2

- Esempio: Data la tabella COMUNE(Nome,Estensione:PG)

  $$ADF_{Estensione.Area,AreaC}(COMUNE)$$

  produce una relazione di schema (Nome,Estensione,AreaC) dove nell’attributo AreaC è presente il risultato dell’applicazione della funzione Area al valore presente nell’attributo Estensione della tupla stessa