Capitolo 12°
Data mining e Warehousing spaziali

Data mining e Warehousing spaziali

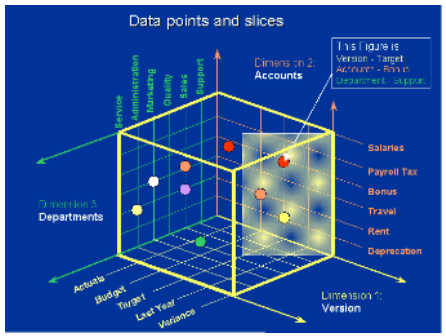
- 12.1 – Riutilizzazione dei vecchi dati
- 12.2 – Datawarehousing
- 12.3 – Data mining
- 12.4 – Data mining spaziale
- 12.5 – Conclusioni

12.1 – Riutilizzazione dei vecchi dati

- Conservare i dati per motivi giuridici dando una struttura comune che consente di fare delle ricerche
  - ➔Datawarehouses (emporio di dati)
- Estrarre il succo dei dati
  - ➔Data mining

12.2 – Datawarehousing

- Archivi storici
- Emporio di dati
- Sistemi OLAP
- Cubi



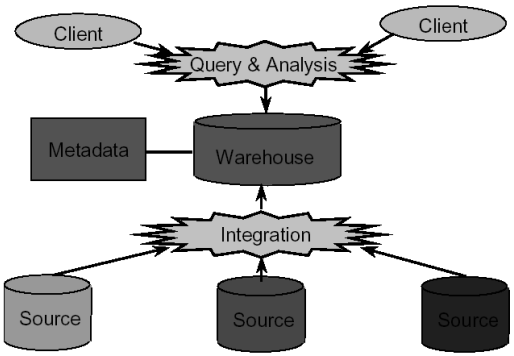
OLAP = Online Analytical Process

What is a Warehouse?

- Collection of tools
  - gathering data
  - cleansing, integrating, ...
  - querying, reporting, analysis
  - data mining
  - monitoring, administering warehouse

CS 245 Notes12 5

Warehouse Architecture



CS 245 Notes12 6

OLTP vs. OLAP

OLTP

- Mostly updates
- Many small transactions
- Mb-Tb of data
- Raw data
- Clerical users
- Up-to-date data
- Consistency, recoverability critical

OLAP

- Mostly reads
- Queries long, complex
- Gb-Tb of data
- Summarized, consolidated data
- Decision-makers, analysts as users

CS 245 Notes12 13

Warehouse Models & Operators

- Data Models
  - relations
  - stars & snowflakes
  - cubes
- Operators
  - slice & dice
  - roll-up, drill down
  - pivoting
  - other

CS 245 Notes12 15

Star

product	prodId	name	price
	p1	bolt	10
	p2	nut	5

store	storeId	city
	c1	nyc
	c2	sfo
	c3	la

sale	oderId	date	custId	prodId	storeId	qty	amt
	o100	1/7/97	53	p1	c1	1	12
	o102	2/7/97	53	p2	c1	2	11
	105	3/8/97	111	p1	c3	5	50

customer	custId	name	address	city
	53	joe	10 main	sfo
	81	fred	12 main	sfo
	111	sally	80 willow	la

CS 245 Notes12 16

Cube

Fact table view:

sale	prodId	storeId	amt
	p1	c1	12
	p2	c1	11
	p1	c3	50
	p2	c2	8

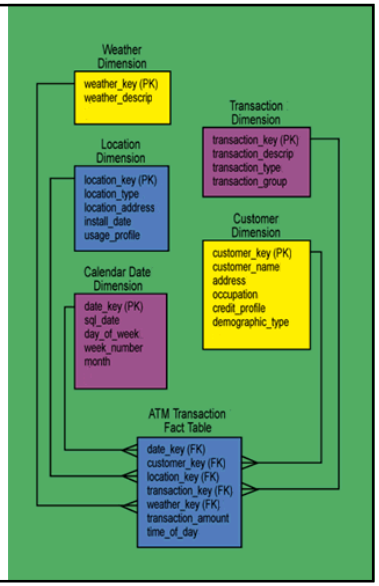
Multi-dimensional cube:

	c1	c2	c3
p1	12		50
p2	11	8	

dimensions = 2

CS 245 Notes12 20

Esempio di struttura



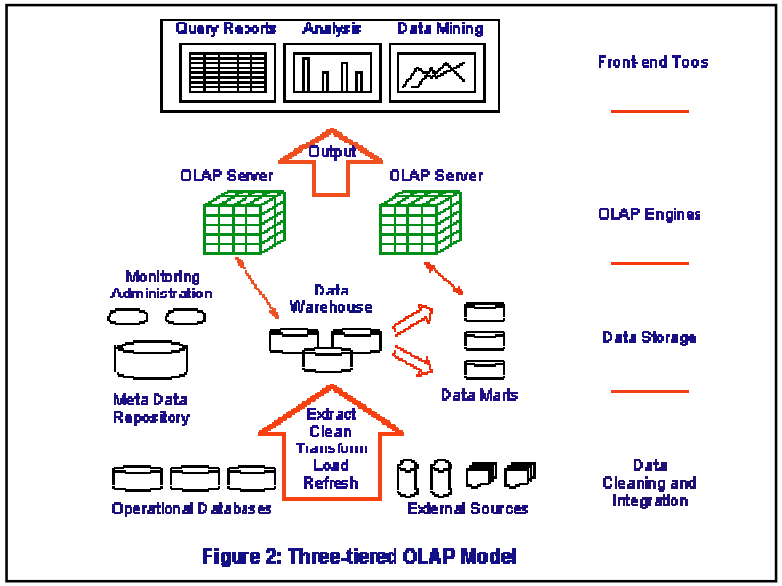
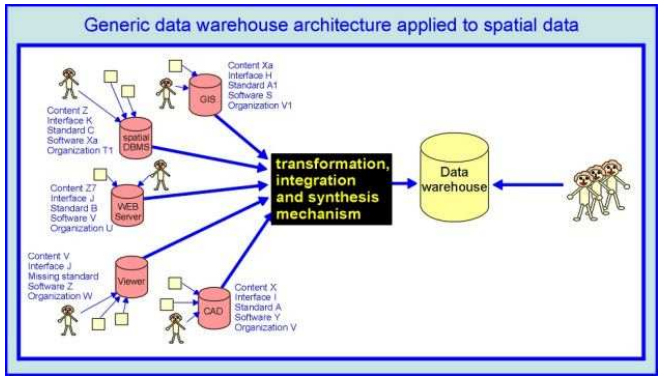
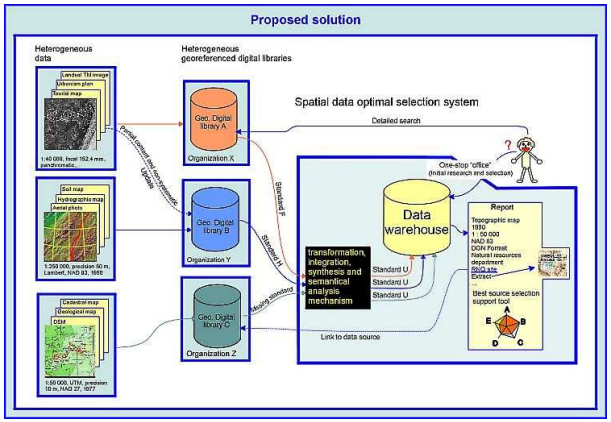


Figure 2: Three-tiered OLAP Model

Datawarehouse spaziale



Datawarehouse spaziale

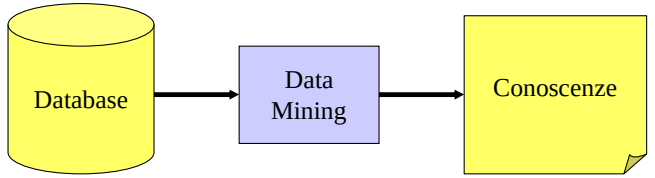


12.3 – Data mining

- Abbiamo tanti dati
- Estrarre il "succo" dei dati
- Metafora:

Database = cava aurifera
Conoscenza = pepite d'oro

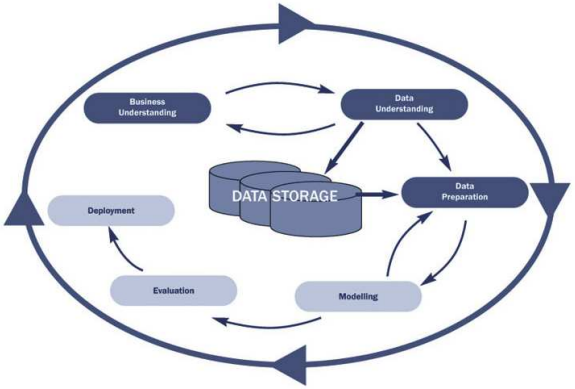
Processo



"Cesto della casalinga"

- Esempio DB relazionale
- *Compere (n°compera, pane, latte, burro, marmellata, insalata, carne)*
- Verificare : *pane* e *latte* ➔ *burro*
- Supporto
- Percentuale di verità

Metodologia generale del Data Mining



12.4 – Data mining spaziale

- Applicazione del data mining agli aspetti spaziali
- Trovare regolarità che possono essere utili nella pianificazione del territorio

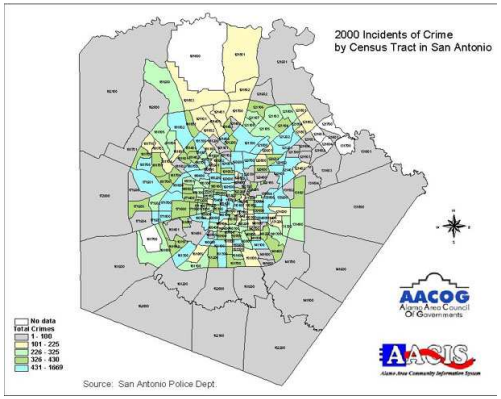
Esempio

- Relazione tra
  - lago
  - strada
  - ristorante

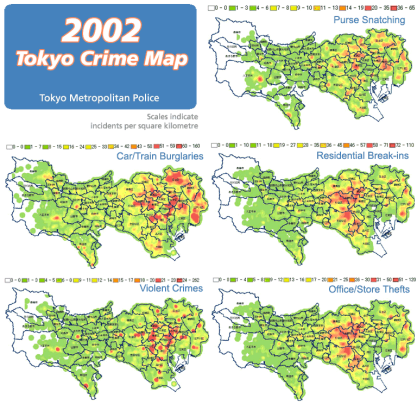
Conoscenza del territorio

- Criminalità
- Incidenti stradali
- Uso del suolo
- Inquinamento
- Economia
- Demografia
- ecc.

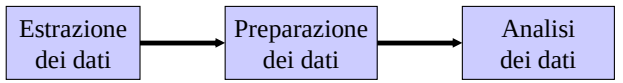
Esempio criminalità



Mappe di criminalità a Tokyo



Fasi principali del Data Mining



Prima legge di Tobler

- "*Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970)"*
- Allora !!!
- Scopo: trovare altre leggi

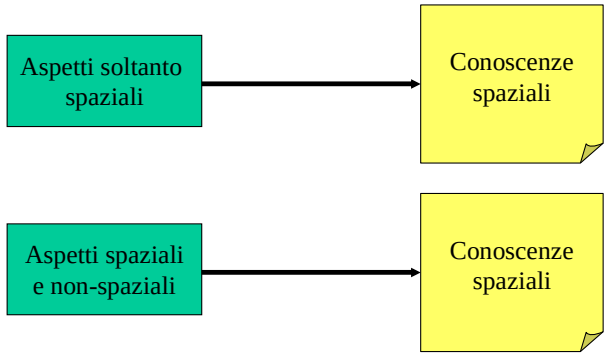
Approcci del Data Mining Spaziale

- Generalizzazione
  - Gerarchia di concetti
- Clustering.
  - Processo di raggruppamento d'oggetti fisici o astratti in classi d'oggetti simili
  - Approcci: Partitioning, Partitioning gerarchico, località, griglie.
- Associazioni spaziali.
  - Regole che descrivano le implicazioni tra diversi caratteristiche, verso un altro gruppo di caratteristiche.

Approcci del Data Mining Spaziale

- Approssimazione ed Aggregazione
  - Perché abbiamo lì una clusterizzazione ?
- Mining in database d'immagini o di raster
  - Può essere visto come un caso particolare.
- Classificazione spaziale
  - Assegnazione di un oggetto a una classe basata sui valori degli attributi
- Scoperta di un trend spaziale
  - Evoluzione regolare di un insieme d'attributi nell'intorno di un altro oggetto

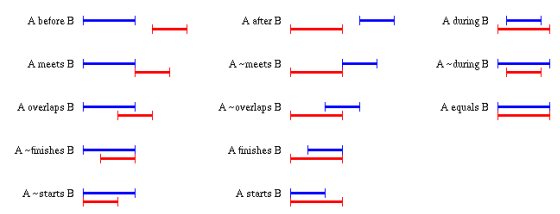
Idee generali



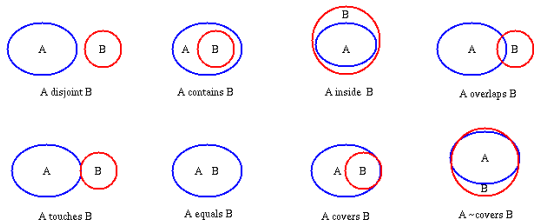
Contesto

- Esistono algoritmi di data mining basati sui grafi
- Rappresentazione grafica
  - dati non-spaziali
  - dati spaziali
  - relazioni spaziali e non-spaziali
- Relazioni spaziali
  - topologiche
  - direzionali
  - distanza
  - mereologiche

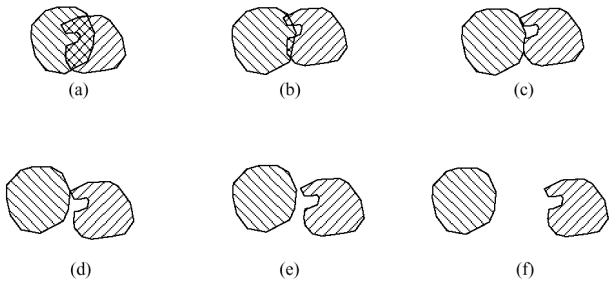
Allen



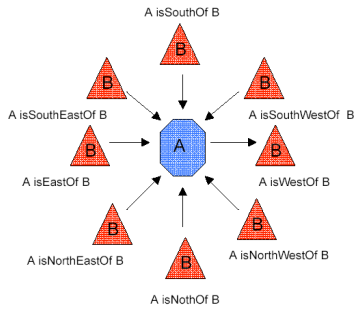
Egenhofer



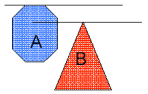
Esempi di relazioni topologiche



Relazioni direzionali



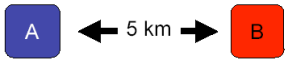
e.g. "*A isNorthOf B*"



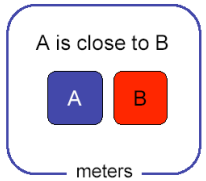
True if northern most point of A is north of the northern most point of B

Relazioni di distanza

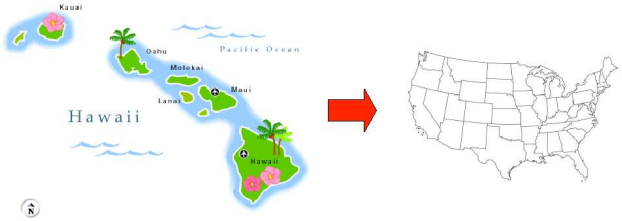
- Quantitative Distance



- Qualitative Distance



Relazioni mereologiche



Hawaii isPartOf United States

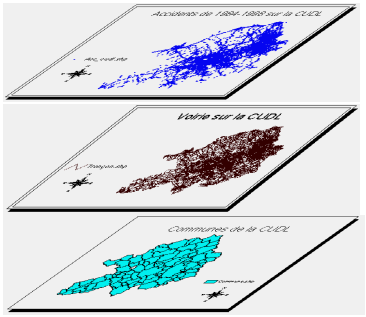
Esempio: incidenti stradali

Incidenti

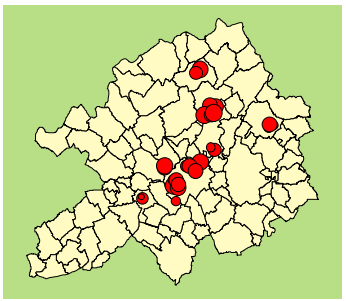
Rete viaria

Comuni

...



Esempio: caratterizzare gli incidenti notturni



Generalizzazione della tabella « incidenti » su due attributi: *giorno* e *luminosità*

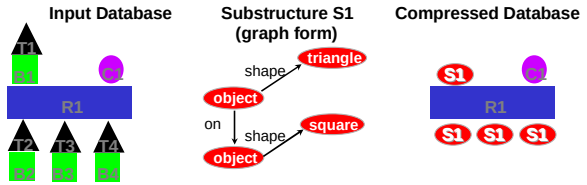
Ricerca delle concentrazioni spaziali locali anormali

Approccio di Pech basato su grafi

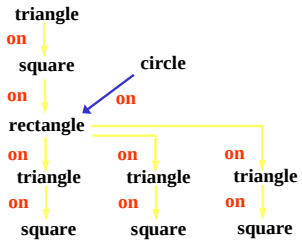
- Trasformazione dei dati in un grafo
  - vertici: oggetti ed attributi
  - spigoli: relazioni tra di loro
- Uso di un algoritmo di ricerca nei grafi
- Non necessario mettere tutti gli oggetti e tutti i loro attributi nel grafo
- Ricerca per pattern particolari
- Uso dell'algoritmo SUBDUE

Algoritmo SUBDUE

- Sviluppato all'Università del Texas
- Ricerca nei grafi
- http://cygnus.uta.edu/subdue



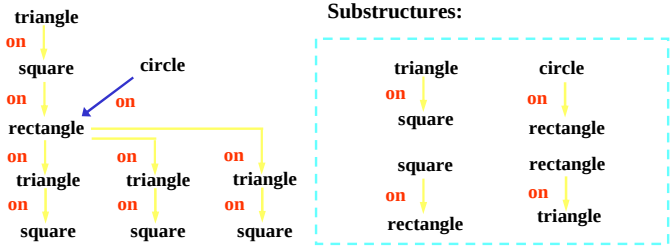
Creare sottostrutture per ogni oggetto



Substructures:

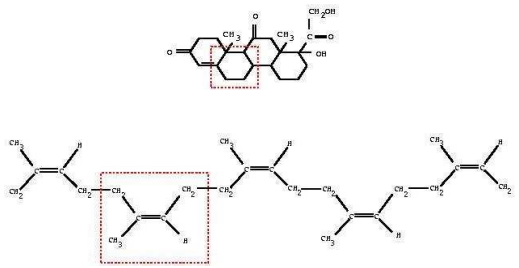
triangle (4), square (4), circle (1), rectangle (1)

Trasformazione

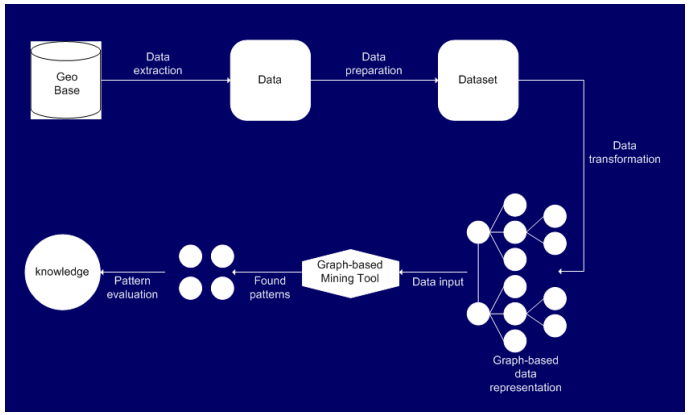


Ricerca della ripetizione di una sotto-struttura

Esempi

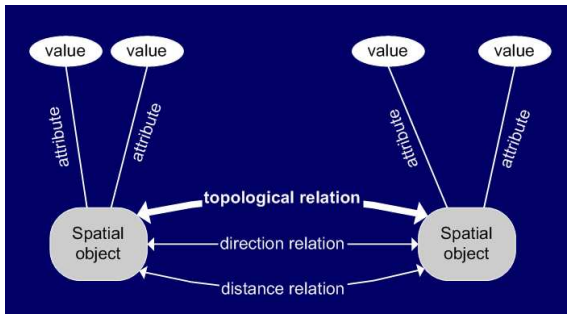


Approccio di Pech

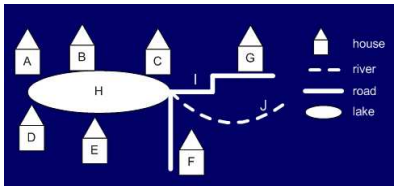


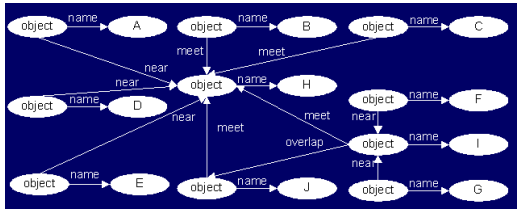
Rappresentazione dei dati

- Vertici: oggetti spaziali ed i loro attributi.
- Spigoli: relazioni spaziali

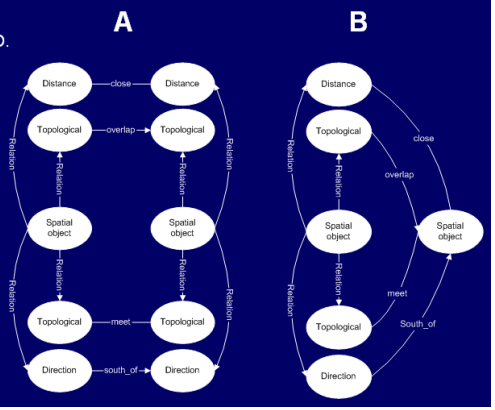


Esempio





Spatial relations:
- Distance: close.
- Topological: overlap.
- Topological: meet.
- Direction: south_of



12.5 – Conclusioni

- Abbiamo tanti database territoriali
- Estrarre il succo dei dati
- Trovare nuove regolarità
- Visualizzazione dei risultati
- Cambio di paradigma
  - passato: scoprire le verità nel mondo reale
  - presente: scoprire le verità nei database