

Graph Databrowsing

Sistemi di analisi e accesso ai dati basati sui grafi

di Ernesto Lastres*, Stefano Corsi

*Sister - Sistemi Territoriali

Until today, those intending to make an analysis of combinations of data could use two approaches: "Data Retrieval", the objective exploration of the data, or "Data Mining", a knowledge-mining process using the application of algorithms to identify "hidden" associations. CINECA, in collaboration with its partner Sistemi Territoriali (Sister), proposes a third approach, called Graph Databrowsing, which permits highlighting of "hidden" relations between data through graphic representation with a simple and effective interface. This means of exploring information and, in particular, the interrelations between the entities under investigation (persons, property, assets, etc.) is finding interesting application in the fight against fraudulent conduct, both criminal and fiscal/administrative (taxes, insurance). This article illustrates the functional principles underlying Graph Databrowsing, including practical examples intended to explain the area of applicability of this information investigation system.

Fino ad oggi chi intende effettuare delle analisi su un insieme di dati dispone di due approcci: il "Data Retrieval", ovvero si cercano conferme attraverso l'esplorazione oggettiva dei dati, o il "Data Mining", un processo di estrazione di conoscenza tramite l'applicazione di algoritmi che individuano le associazioni "nascoste" tra i dati.

Il Dipartimento di Business Intelligence di CINECA, in collaborazione con il partner Sistemi Territoriali (Sister) propone un terzo approccio, denominato Graph Databrowsing, che consente di evidenziare le relazioni "nascoste" tra i dati mediante una rappresentazione grafica caratterizzata da un'interfaccia semplice ed efficace.

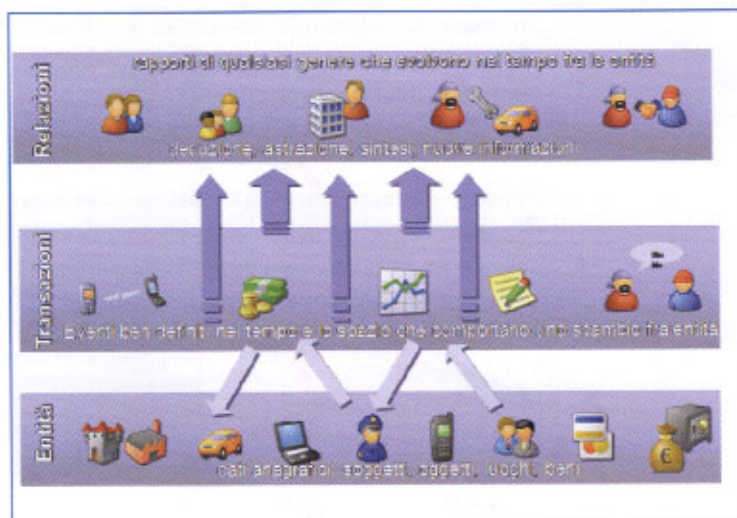
Questa modalità di esplorazione delle informazioni e, in particolare, delle interrelazioni tra le entità oggetto di indagine (persone, proprietà, beni, ecc.) sta trovando interessanti applicazioni nella lotta ai comportamenti fraudolenti, siano essi di natura criminale che di tipi fisca-

le/amministrativo (fisco, assicurazioni).

Il Graph Databrowsing è una tecnologia che fonde insieme tecniche di navigazione ed esplorazione di dati multidimensionali con concetti della teoria dei grafi, per scoprire collegamenti impliciti fra dati logicamente scon-

Graph Databrowsing
Data analysis and access systems based on graphs

Schema architetturale di base del sistema



in base alle proprie esigenze. In questo modo si potrebbero scoprire velocemente collegamenti impliciti fra gli oggetti che altrimenti sfuggirebbero all'attenzione dell'operatore.

Ontologie: analisi semantica, valutazione e controllo dell'informazione. Una ontologia è la rappresentazione di un insieme di concetti in un dominio ben definito e delle relazioni fra di essi. Le ontologie possono aiutarci in questo modo a definire il significato dei dati che trattiamo e di conseguenza ci permettono il ragionamento automatizzato su di essi. Le ontologie sono strettamente legate al concetto di grafo, sia in termini di definizione dei concetti e relazioni che in termini del trattamento dell'informazione.

Il Modello

Il modello del Graph Databrowsing è suddiviso in tre livelli strettamente collegati:

Livello Entità

È costituito da tutti gli Attori materiali o non, che fanno parte del mondo che vogliamo modellare. Un attore/entità è un esecutore di azioni e di solito ha un ruolo all'interno del nostro modello. Esempi di attori possono essere: una persona fisica o giuridica, un telefono cellulare, una carta di credito, un conto in banca, una compagnia assicuratrice, una polizza, ecc. Le entità hanno uno stato che può variare nel tempo e quindi possono avere una storia, e possono essere strutture dati complesse con relazioni statiche fra di loro (ad esempio: Persona ↔ Carta di Identità).

Livello Transazionale

È costituito da eventi ben definiti nel tempo e lo spazio e che coinvolgono una o più entità (attori) del nostro modello. Questi eventi spesso comportano lo scambio di informazione fra entità (da cui evento transazionale). Possono essere eventi: una telefonata, un bonifico, un pagamento con Carta di Credito, un acquisto, un sinistro, un furto, un attentato, ecc.

Livello Relazionale

Questo è un livello di sintesi che permette di modellare i collegamenti di diverso tipo che esistono fra le entità (attori) del nostro modello. Per questo motivo viene visto come una sintesi di quello che accade nel mondo reale

(sintesi del livello transazionale). Queste relazioni possono essere derivate a partire dagli attributi degli attori stessi e dalle informazioni registrate nel livello transazionale. Diversamente dalle transazioni, le relazioni non sono ben definite nel tempo e nello spazio, semplicemente possono avere un intervallo temporale di validità. Le relazioni possono essere informazioni di prima mano oppure possono essere derivate a partire dalla informazione di base (entità e transazioni). Esempi di relazioni possono essere: è figlio di, è amico di, è proprietario di, è usato da, è socio di, è conoscente di, è coniuge di, è capo di, è impiegato di, ecc.

Il Grafo

In quest'ottica, il grafo è costituito da entità come nodi del grafo, e da relazioni come archi di esso. Gli eventi e transazioni sono dettagli riferiti agli archi del grafo.

L'implementazione

Astrazione

Per implementare il nostro modello è necessario modellare secondo questa filosofia (entità - eventi - relazioni), astrando e semplificando gli attributi per concentrarci sui problemi da risolvere. Nel fare ciò buona parte dei dati di business vengono esclusi o semplificati.

ETL (Extraction Transformation and Loading)

Per questo motivo è necessario un meccanismo di estrazione di informazione, trasformazione e caricamento che prenda i dati di business di interesse, li trasformi e semplifichi nel modo adeguato e poi li carichi nel nostro modello. L'estrazione dovrebbe avvenire in automatico a partire dai Datawarehouse di business e da informazioni aggiuntive fornite in supporti vari e con mezzi vari.

Derivazione

Una volta popolato il nostro modello possiamo eseguire processi di derivazione automatici più inerenti al nostro modello di Graph Databrowsing che permettono di determinare relazioni implicite nei dati.

Lo strumento di analisi interattiva

Permette di visualizzare il grafo (entità e relazioni) in modalità grafica, con la possibilità di

