

Università degli Studi di Salerno
Dipartimento di Scienze Aziendali - Management & Innovation System

Dottorato di Ricerca in Management and Information Technology
Curriculum Informatica, Sistemi Informativi e Tecnologie del Software
XVI Ciclo



Ph.D. Thesis

**Information Visualization:
from Petroglyphs to CoDe Graphs**

Candidate

Dr. Paola de Roberto

Coordinator

Prof. Andrea De Lucia

Tutor

Prof. Genoveffa Tortora

2016 - 2017

Abstract

La visualizzazione dei dati riguarda la comunicazione dei dati attraverso rappresentazioni e tecniche visuali. Essa mira a migliorare la percezione e supportare il processo decisionale basato sui dati, in modo da ottenere approfondimenti altrimenti difficili da raggiungere. Una buona visualizzazione dei dati consente di identificare i modelli e di comprendere meglio i fenomeni. In altre parole, la visualizzazione dei dati è correlata a un'innata capacità umana di comprendere rapidamente, discernere e convertire schemi in informazioni utili e utilizzabili.

Già nel 35.000 a.C gli essere umani usavano rappresentazioni grafiche visuali, utilizzando le incisioni rupestri. Infatti, già i nostri antenati ragionavano in termini di **modelli** o **schemi**: la rappresentazione visuale dell'informazione è un concetto antico, come testimoniano le incisioni rupestri ritrovate. Nel corso dei secoli, la visualizzazione delle informazioni si è evoluta per tenere conto delle mutevoli esigenze umane e il suo utilizzo è diventato sempre più consapevole. Le prime tecniche di visualizzazione dei dati sono state sviluppate per osservare e rappresentare quantità fisiche e posizioni geografiche e di oggetti celesti.

Successivamente, l'uso combinato della geometria euclidea e dell'algebra ha migliorato l'accuratezza e la complessità della rappresentazione dell'informazione, in diversi campi come l'astronomia, la fisica e l'ingegneria. Infine, nel secolo scorso sono state elaborate forme più moderne di rappresentazione di dati: a partire da diagrammi, istogrammi e grafici fino ad arrivare a dati multidimensionali e visualizzazioni dinamiche e interattive di dati temporali [41].

Oggi, l'enorme quantità di informazioni consente un'interpretazione più precisa dei fenomeni, favorendo così l'adozione di tecniche di infografica, in particolare per supportare il processo decisionale manageriale nel settore aziendale.

In realtà, la visualizzazione dei dati rappresenta il cuore della *Business Analytics*, che ha bisogno di trasformare una grande quantità di dati complessi in informazioni comprensibili per una descrizione più intuitiva dei fenomeni.

La visualizzazione delle informazioni utilizza *report* visuali che possono essere considerati *sentenze* di un linguaggio visuale in uno specifico dominio. E' necessario quindi, esaminare e affrontare i principali problemi relativi ai linguaggi visuali e alla visualizzazione delle informazioni in diversi ambiti: i) attraverso l'analisi e l'interpretazione delle informazioni visuali; ii) studiando l'uso della modellazione delle informazioni e le tecniche di rappresentazione necessarie per

progettare e implementare sistemi di supporto alle decisioni incentrati sull'utente.

Nella prima parte della tesi, proponiamo un sistema di *Visual Analytics* applicato all'ambito archeologico; questo sistema supporta gli archeologi rupestri nell'analisi e classificazione di migliaia di rilievi e manufatti; ogni artefatto contiene diverse incisioni, chiamate petroglifi.

L'arte rupestre è un termine coniato nell'archeologia per indicare qualsiasi segno umano inciso sulla pietra naturale [19,65]. La maggior parte dei simboli riguardanti l'arte rupestre sono rappresentati da petroglifi, che sono stati creati rimuovendo parte di una superficie rocciosa incidendo, raccogliendo, intagliando e abradendo. Sebbene non sia possibile dare certe interpretazioni a questi petroglifi, gli archeologi hanno proposto molte teorie per spiegare il loro significato, ad esempio astronomico, culturale o religioso [105]. A tal fine, presentiamo un sistema di analisi visuale, denominato DARK, per supportare gli archeologi di arte rupestre nell'esplorazione di depositi di scene di arte rupestre, ciascuna costituita da centinaia di petroglifi scolpiti da antichi popoli sulle rocce. Data la loro crescente complessità, analizzare questi *repository* di informazioni eterogenee è diventato un compito ed una sfida importanti per gli archeologi di arte rupestre. DARK combina tecniche di visualizzazione con analisi fuzzy di scene di arte rupestre per dedurre informazioni cruciali per la corretta interpretazione delle scene. Inoltre, le viste DARK consentono agli archeologi di convalidare la loro ipotesi rispetto alle informazioni memorizzate nei repository. Inoltre, descriviamo e dettagliamo le caratteristiche principali dell'applicazione *mobile* PetroSketch per supportare gli archeologi nella classificazione e nel riconoscimento dei petroglifi.

PetroSketch è un taccuino virtuale che consente agli utenti di disegnare un petroglifo su una pagina bianca o di seguirne il contorno dopo averlo acquisito con la fotocamera in modo da classificarlo e ottenere l'elenco dei simboli più simili ad esso. Quest'ultima azione viene eseguita da un algoritmo flessibile di *image matching* che misura la somiglianza tra i petroglifi usando una distanza, derivata dal modello di deformazione dell'immagine, che è computazionalmente efficiente e robusta alle distorsioni locali.

Nella seconda parte della tesi, focalizziamo la nostra attenzione sullo studio del problema della visualizzazione dei dati nell'ambito della *Business Analytics*. In questo contesto, viene affrontato il problema della reportistica visuale di informazioni estratte da sorgenti di dati e in particolare da data warehouse. Gli strumenti di visualizzazione giocano un ruolo centrale in contesti in cui le informazioni devono essere rappresentate preservando l'accuratezza dei dati e la complessità delle relazioni tra di essi. Molta attenzione è stata dedicata al problema della visualizzazione efficace dei dati nei singoli *report* che di solito vengono visualizzati attraverso diversi tipi di grafici standard

(istogrammi, grafici a torte, ecc.) o in forma tabellare. Tuttavia, questo tipo di rappresentazione fornisce item informativi separati ma non fornisce alcun supporto per visualizzarne le relazioni esistenti tra di essi che sono alla base della maggior parte dei processi decisionali. I *decision maker* impiegano molto tempo e sforzo nell'interpretazione di grafici derivati da grandi database multidimensionali. Quindi, la scelta di una rappresentazione grafica è fondamentale ogni volta che è necessario interpretare i dati. I dati sono solitamente rappresentati da diversi diagrammi, come istogrammi e torte, che non evidenziano le relazioni logiche tra di loro.

Il linguaggio CoDe (Complexity Design) consente di organizzare visualizzazioni, denominate Grafici CoDe, componendo e aggregando graficamente diagrammi attraverso collegamenti concettuali. La modellazione della composizione del grafico basata su CoDe consente di visualizzare le relazioni tra le informazioni nella stessa immagine in base alla definizione di efficienza di una visualizzazione fornita da Jacques Bertin [9]: "La costruzione (grafica) più efficiente è quella in cui ad ogni domanda, qualunque sia il suo tipo e livello, può essere data risposta in un singolo istante di percezione, cioè in una singola immagine ". Questa rappresentazione denominata modello CoDe può essere considerata una mappa cognitiva di alto livello delle complesse informazioni contenute nei dati di base. La scelta del *layout* di visualizzazione finale in termini di grafici standard viene affidata a un'interfaccia di visualizzazione che fornisce i costrutti di implementazione necessari.

L'applicazione della metodologia CoDe incide principalmente nel campo della *Business Analytics* in cui la conoscenza, la gestione e l'analisi dei dati aziendali (ad es. vendite, produzione, costi e profitti) sono requisiti fondamentali per un valido sistema decisionale. Infatti, nelle aziende moderne, l'informazione strategica è necessaria per garantire la solidità e la sopravvivenza delle imprese. Inoltre, CoDe è un modello grafico alternativo per la rappresentazione dei dati che riduce il divario tra l'esperto di data warehouse e il manager aziendale. Infatti, CoDe consente di selezionare le informazioni di interesse, modellare le relazioni tra di esse e generare automaticamente i diagrammi contenenti queste informazioni collegati attraverso le relazioni individuate. A tal fine, presentiamo il processo di generazione dei CoDe Graph e, dopo aver analizzato lo stato dell'arte relativo alla valutazione della comprensibilità della rappresentazione grafica, proponiamo innanzitutto una classificazione di tali approcci valutativi e quindi eseguiamo la valutazione della comprensibilità dei CoDe Graph rispetto ai report tradizionali mediante un esperimento controllato, che ha coinvolto 47 partecipanti. I risultati mostrano che i CoDe Graph riducono lo sforzo dei partecipanti, migliorando al contempo l'efficacia e l'efficienza in termini di comprensione, dimostrando l'utilità della metodologia CoDe.