

P ROGETTAZIONE DI BASI DI DATI E SISTEMI INFORMATIVI

Lucio Colizzi

Divisione Ingegneria Informatica

Consorzio CETMA (Centro Progettazione Design & Tecnologie dei Materiali)

Introduzione

Il corso di *Progettazione di basi di dati e sistemi informativi* è stato configurato per venire incontro alle specifiche esigenze di chi, avendo utilizzato diversi strumenti a fini diagnostici e conservativi, si propone di individuare un metodo rigoroso per la gestione della grossa mole di dati ottenuti. L'obiettivo di un Sistema Informativo è in questo caso quello di proporre meccanismi (il più possibile automatici) per l'estrazione di informazioni, in modo aggregato, e che tenga in giusto conto della molteplicità ed eterogeneità delle varie sorgenti di dati, siano essi strumentali (analisi UV, fluorescenza, termografia, scansione 3D, analisi chimico/fisiche, etc.) o puramente ottenuti da studi di natura storico/tecnica, antropologica, iconografica, stilistica, etc. Il percorso così costruito permette di acquisire quelle conoscenze di base relative alla modellazione e strutturazione dei dati e la relativa traduzione fisica in sistema informatico.

Sistema Informativo e Sistema Informatico

Cominciamo con il sottolineare la distinzione, spesso poco nota, tra i termini Sistema Informativo e Sistema Informatico. Mentre il primo attiene alla componente di una organizzazione che gestisce, acquisisce, elabora e conserva le informazioni di interesse (ovvero quelle utilizzate per il perseguimento degli scopi dell'organizzazione); il secondo è costituito dalla porzione automatizzata del sistema informativo, ovvero quella parte dello stesso, che gestisce le informazioni mediante tecnologie informatiche. Da questa distinzione deriva il fatto che il SI esiste come conseguenza diretta dell'esistenza di una organizzazione che produce dati e informazioni. Ad esempio, una biblioteca, un archivio documentale cartaceo e/o membranaceo, etc. sono tutti esempi di sistemi informativi ora come lo erano nei tempi antichi. Mentre le attuali tecnologie informatiche di gestione della conoscenza permettono di realizzare sistemi informatici utili all'elaborazione complessa dei dati al fine di porre in evidenza informazioni utili all'analisi storico-tecnica.

L'approccio generale nello studio di un sistema informativo/co per la gestione delle raccolte e collezioni in archivio e biblioteca e trattamento dei flussi documentali, è ormai consolidato nell'ambito dell'ingegneria del software in diverse fasi:

- Studio di fattibilità
- Specifica dei requisiti
- Progettazione esecutiva
- Realizzazione
- Funzionamento
- Manutenzione

Ma il pilastro fondamentale per la buona riuscita di un sistema informatico/vo è la modalità con cui si descrive il dominio della conoscenza relativo al contesto applicativo, e conseguentemente la struttura dei dati, sui quali le applicazioni si interfacciano per estrapolare informazioni. L'argomento va anche esso affrontato con idonee metodologie che portano alla definizione di modelli logici e fisici.

Dato e Informazione

A questo punto è utile vedere meglio il concetto di dato e informazione. Il primo può essere definito come *mattoncino* elementare di informazione, immediatamente presente alla conoscenza, prima di qualsiasi elaborazione e/o interpretazione. Il dato è ciò che descrive, in termini conoscitivi un oggetto, una realtà o un contesto. L'informazione è il valore aggiunto, in termini conoscitivi, estraibile dai dati attraverso attività di sintesi, aggregazione, interpretazione e proiezione. Pertanto il dato è un elemento della conoscenza meramente descrittivo ed attiene a componenti che potremmo definire statiche (generalmente invarianti nel tempo). L'*informazione* è il risultato di un processo elaborativo ed interpretativo, e per questo riguarda le componenti *dinamiche* di un sistema informativo/co. Quando si parla di *database*, ci si riferisce sempre ed in ogni caso alla gestione delle componenti statiche (ovvero dei dati); mentre l'estrazione di informazioni utili all'utente è effettuato da applicazioni studiate e sviluppate ad-hoc.

Da quanto detto si comprende come un efficace e corretto studio dei dati in gioco, in un determinato contesto applicativo, pone i pilastri per un sistema informativo/co efficace (in grado di rispondere alle esigenze dell'utenza) ed efficiente.

La fase di analisi dei dati è oggi supportata da metodologie complete che si rifanno a standard internazionali ormai universalmente accettati, ovvero il modello ER (Entità Relationship Model) ed il Modello Relazione o delle tabelle [1]. Il primo permette una modellazione concettuale dei dati in gioco, mentre il secondo traduce il modello concet-

tuale in una serie di tabelle correlate da implementare direttamente all'interno di un DBMS (Data Base Management System).

La definizione dei modelli precedentemente citati parte dall'individuazione delle Entità in gioco e da come queste entità sono correlate tra di loro.

Un'entità è definibile come una classe di oggetti (fatti, persone, cose) appartenenti al mondo che si intende modellare con caratteristiche di autoconsistenza. Una relazione esprime il grado e la modalità di correlazione tra due o più entità.

Un esempio potrebbe essere il seguente:



in cui sono state individuate due entità AFFRESCO e PIGMENTO ed una relazione. Un affresco presenta m tipi di pigmenti, ma lo stesso tipo di pigmento può essere ritrovato in più affreschi nella base dati. Per sapere quali pigmenti sono riscontrati in un affresco specifico sarà sufficiente effettuare un *prodotto cartesiano* tra le istanze di AFFRESCO e PIGMENTO collegate dalla relazione "presenta".

Utilizzando questi concetti è possibile definire modelli di dati per qualsiasi contesto applicativo indipendentemente dalla complessità. Vediamo di seguito un possibile scenario .

Un esempio concreto

Dopo uno studio approfondito delle tecniche di ingegneria software per la strutturazione dell'informazione, si è affrontato con gli studenti del corso un caso concreto. La discussione e l'analisi di gruppo (di fondamentale importanza nella progettazione delle basi di dati) ha portato ad una interessante struttura di dati. Lo scenario prospettato è stato il seguente:

"La soprintendenza di XXX vi vuole affidare un lavoro di catalogazione di innumerevoli opere situate in un museo/biblioteca. Le opere sono state accatastate negli anni, alla rinfusa in ogni angolo dello stabile. Da una prima indagine sul posto, si è potuto riscontrare che esistono nella collezione diverse tipologie di opere tra cui, libri e manoscritti, dipinti, tele ed opere lignee, anche la struttura risulta affrescata in più parti e sul pavimento sono stati creati diversi mosaici. In diversi secchi sono stati accatastati sotto forma di rotoli innumerevoli documenti (atti, contratti, testamenti, etc.). La situazione di preca-

rietà con cui si sono conservate le opere ha generato diversi processi di degrado (lacune, muffe, viraggi cromatici, attacchi biologici, etc.). La soprintendenza vorrebbe catalogare tutte le opere presenti nel sito attraverso un lavoro sistematico, e nel prossimo futuro vorrebbe che si sviluppasse un vero e proprio sistema informativo, che oltre a contenere informazioni sulle opere tenga traccia anche della loro situazione di degrado, in modo da attuare politiche di intervento mirate.”

La discussione in aula ha portato ad una definizione (ovviamente non l'unica possibile!) di un modello dei dati di tipo EER, che è in grado di rispondere ai quesiti apparentemente complessi posti dal caso sopra citato.

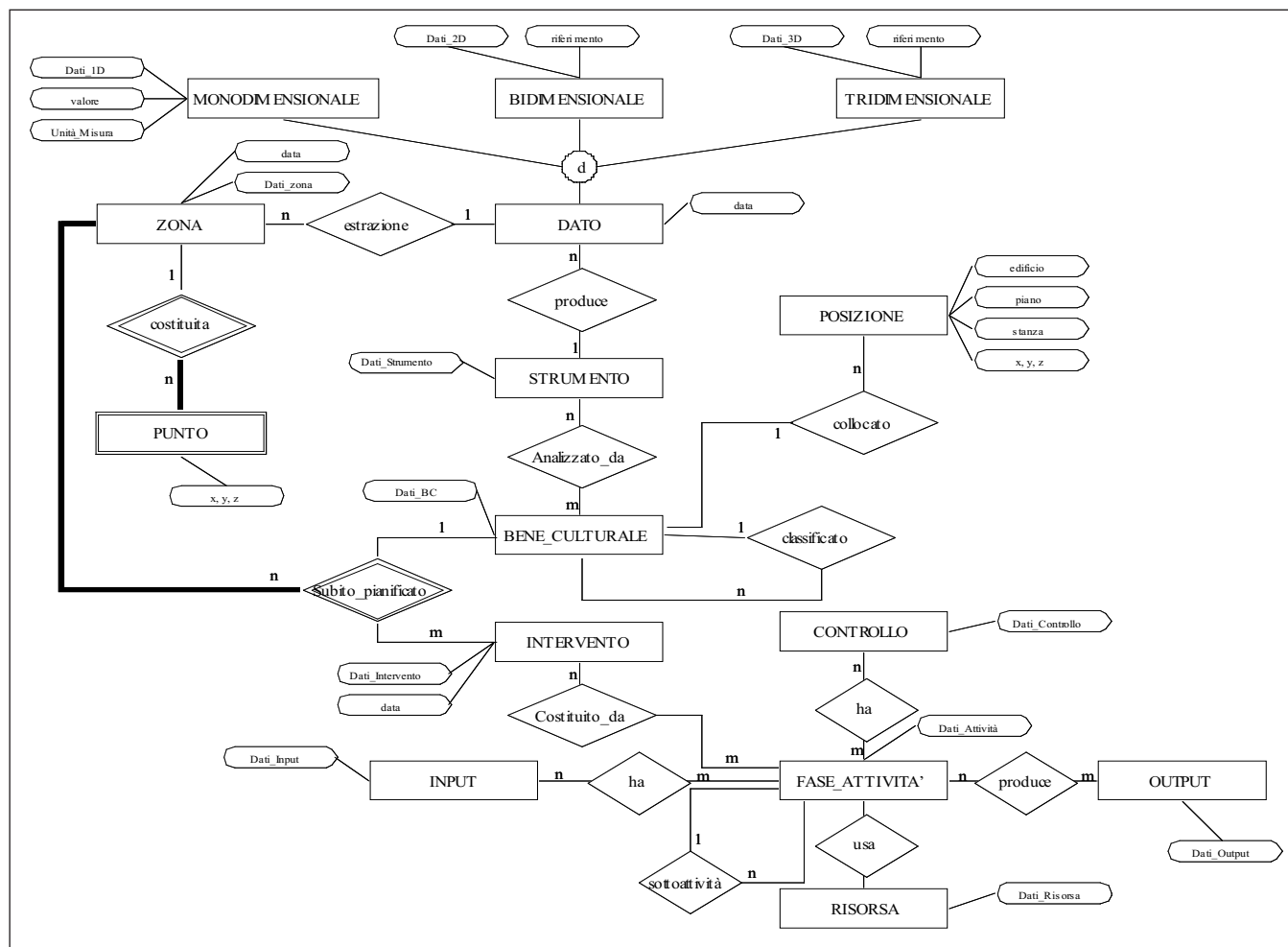
Per capire le potenzialità del modello progettato in aula, possiamo osservare che l'entità principale individuata è l'opera o "BENE_CULTURALE": questa ha una autorelazione che rappresenta un livello multiplo di classificazione del bene (es. opere lignee, bronzi, affreschi, libri e manoscritti, mosaici, manufatti, etc.). Ogni bene può essere analizzato da uno o più strumenti ed ogni strumento può fornire uno o più dataset (es. immagine nel visibile, fluorescenza UV, spettro di massa, infrarosso, termografia, etc.). La base dati effettua un distinguo nel caso di dataset monodimensionali (es. concentrazione di una molecola in un punto o campione), bidimensionali (immagini a varie frequenze) o tridimensionali (stratigrafie, scansioni laser 3D, composizioni volumetriche etc.). Ogni bene è relazionato con la propria anamnesi, ovvero la storia degli interventi passati, ed ogni intervento è visto come un processo ben definito. La modalità di rappresentazione del processo segue le indicazioni dello standard IDEF0 [2] che vede un processo P, come una procedura che trasforma un set ben definito di input in un set definito di output, utilizzando n risorse, sottoposto ad un insieme di n vincoli. L'utilizzo di strumenti di diagnostica, ed osservazioni dirette sul bene culturale, consente di individuare delle ZONE di degrado, delimitate e definite da un insieme di PUNTI che rappresentano i vertici di una spezzata che circoscrive la zona di interesse. Ogni zona è monitorata nel tempo e consente di evidenziare l'andamento del degrado nel tempo.

Di seguito riportiamo il modello Entità Relazioni Avanzato (tab. 1).

Dall'analisi del modello emergono le sue potenzialità. In linea di massima possiamo evidenziare le interrogazioni in termini di estrazione di informazioni, che la base dati è in grado di fornire:

- Quali sono le informazioni su un BC custodito, dove è posizionato e quale è stata la storia delle sue localizzazioni?
- Con quali strumenti è stato analizzato a fini diagnostici?
- Quali sono stati gli esiti dell'analisi?

Tabella 1. Modello Entità Relazioni Avanzato.



- Esistono delle zone di degrado? Qual' è la loro dimensione?
- Come si evolvono nel tempo le zone di degrado?
- Sono stati effettuati interventi sull'opera? Quali interventi?
- Quali fasi e attività sono state effettuate durante un intervento e quali zone del bene sono state interessate?
- Quali sono gli interventi pianificati per il futuro?

Ovviamente a queste informazioni direttamente ricavate dalla base dati è possibile individuarne altre ottenute da elaborazione dei dati di base. Esempi potrebbero essere i seguenti:

- *“Qual è il prossimo intervento da effettuare tenendo conto dell'evoluzione nel tempo del degrado?”*
- *“Come posso confrontare nel tempo e nello spazio misure provenienti da strumenti differenti?”*

Una risposta a questa domanda potrebbe essere data a livello applicativo attuando politiche di controllo nel tempo dei processi di degrado che coinvolgono zone ben definite.

Se dovessimo collocare i dati riportati nell'interfaccia (fig. 1) nell'ambito della nostra base dati, possiamo dire che l'immagine del Cristo è stata acquisita con uno spettrofotometro (gestito dall'entità STRUMENTO) operante nella regione del visibile. L'immagine è un dataset 2D, e sarà inserita nella parte di database relativa all'entità composta DATO –

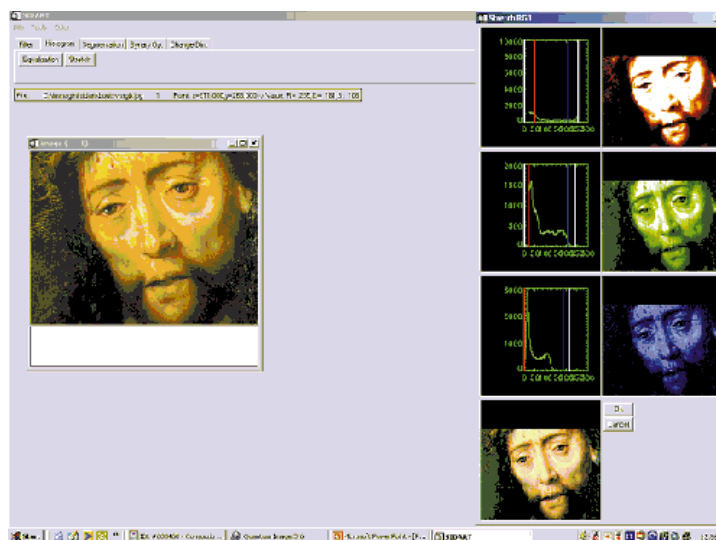


Figura 1. Una possibile interfaccia del sistema informativo.

BIDIMENSIONALE. Allo stesso modo saranno trattate le tre componenti principali in alto a destra, ottenute dalla separazione del Rosso, del Verde e del Blue. Accanto alle tre componenti RGB, osserviamo che è stato riportato il relativo istogramma che però non è presente nella nostra base dati poiché è ottenuto da una elaborazione *al-volo* delle componenti stesse.

Conclusioni

La bontà di una base di dati è valutata sulla base delle potenzialità di estrazione di informazioni che la stessa consente. Ovviamente un sistema informativo ben strutturato è un valido ausilio alle attività di archiviazione e recupero di dati utili alla definizione di quadri diagnostici, ma ancora oggi non sostituisce l'attività dell'esperto che, sulla base di una specifica problematica, deve prendere delle decisioni, dalle quali dipende il futuro dell'opera. Un approccio di gruppo multidisciplinare rappresenta oggi un *modus operandi* vincente in termini di efficacia ed efficienza.

Bibliografia

- [1] ATZENI P., CERI S., PARABOSCHI P., TORLONE S. 1999, *Database Systems* (2nd ed.), Milano, McGraw-Hill.
- [2] BRACCHI G., MOTTA G. 1993, *Progetto di sistemi informativi*, Milano, ETAS Libri.
- [3] RAMEZ ELMASRI, SHAMKANT B. NAVATHE 2000, *Fundamentals of Database Systems* (3th ed.), Redwood City (CA), The Benjamin/Cummings Publishing Company.

Riassunto

Si intende far presente che un approccio sistematico nella progettazione di una base di dati consente di trattare informazioni apparentemente complesse in modo esaustivo, efficiente ed efficace. Le metodologie di riferimento sono ormai consolidate nell'ambito dell'ingegneria del software e trovano applicazione diretta in diversi contesti applicativi, dove il processo di gestione, aggregazione ed analisi di grosse moli di dati risulta critico. Il modello di base per l'analisi logica delle strutture e l'insieme dei costrutti utili alla rappresentazione dei dati si rifanno al cosiddetto Modello Relazionale. La filosofia di estrazione delle informazioni è ricondotta alla manipolazione di Tabelle contenenti istanze di dati ed a semplici prodotti cartesiani.

Abstract

A systematic approach in the project of a database allows to treat information, apparently complex, in an exhaustive, efficient and effective way. The reference methodologies are now consolidated in the field of software engineering and they found direct application in different applicative contexts where the managing and aggregation processes and the analyses of big sets of data are critic. The base model for the logic analysis of the structures and the set of the constructions useful to data representations refer to the so-called Relational Model. The philosophy of information extraction is re-conducted to the manipulation of Tables containing data instances and to simple Cartesian products.

Résumé

Une approche systématique dans le projet d'une base de données consent de traiter des informations apparemment complexes de façon exhaustive, efficiente et efficace. Les méthodologies de référence sont désormais consolidées dans le domaine de l'ingénierie du logiciel et trouvent une application directe dans divers contextes d'application, où le procès de gestion, d'agrégation et d'analyse de grosses masses de données résulte être critique. Le modèle de base pour l'analyse logique des structures et l'ensemble des sens utiles à la représentation des données se réfèrent au soi-disant Modèle Relationnel. La philosophie d'extraction des informations est reconduite à la manipulation de Tableaux contenant des instances de données et à de simples produits cartésiens.

Zusammenfassung

Eine systematische Haltung in der Vorbereitung einer Datenbank ermöglicht es, scheinbar komplexe Informationen auf eine komplette, leistungsfähige und wirksame Art und Weise zu behandeln. Die Bezugsmethodologien sind seit langem im Rahmen des Software-Engineering bekannt und finden direkte Anwendung in verschiedenen Situationen, in denen der Prozess der Verwaltung, Sammlung und Analyse von einer großen Menge von Daten kritisch ist. Das grundlegende Modell für die logische Analyse der Strukturen und die gesamten Systeme, die zur Darstellung der Daten dienen, stützen sich auf dem sogenannten Verbindungsmodell. Die Philosophie der Entnahme der Informationen basiert auf der Verwendung von Tabellen mit Daten und von einfachen kartesischen Produkten.

Resumen

En este artículo vamos a ver que un enfoque sistemático en el proyecto de una base de datos, consiente tratar informaciones aparentemente complejas de manera exhaustiva, eficiente y eficaz. Las metodologías de referencia, ya resultan consolidadas en el ámbito de la ingeniería del software y hallan aplicación directa en diferentes contextos para los cuales el proceso de gestión, agregación y análisis de grandes cantidades de datos resulta ser crítico. El modelo de base para el análisis lógico de las estructuras y el conjunto de las construcciones útiles para la representación de los datos remontan al así llamado Modelo Relacional. La filosofía de extracción de las informaciones remonta a la manipulación de Tablas que contienen instancias de datos y a simples productos cartesianos.

резюме

В статье показывается, как систематический подход в проектировании базы данных, позволяет обрабатывать кажущуюся сложной информацию эффективным, действующим и исчерпывающим образом. Методологии относительно аргумента уже упрочились в инженерной области и находят прямое применение в различных прикладных контекстах, где процесс управления, соединения и анализа масштабного количества данных оказывается критическим. Базовой моделью для логического анализа структур или в целом конструкций полезных для представления данных становится так называемая Относительная модель. Философия извлечения информации сводится к манипулированию Таблицами, содержащими инстанциями данных и к простому картезианскому продукту.