

Informazione e Sistemi Informativi

Contenuti parzialmente tratti da materiale del Prof. Maggini e del Prof. Pedreschi

Sommario

- ◆ Sistemi di gestione dell'informazione
 - Caratteristiche generali
 - Classificazione
- ◆ Tipi di informazione
 - Dati numerici, dati testuali, dati multimediali
 - Documenti, Oggetti, Tuple e Basi di Dati
- ◆ Struttura e contenuto
 - dati strutturati, non strutturati e semi-strutturati
 - modelli dei dati

2

Sistema Informativo (SI)

- ◆ "Sistema"
 - insieme organizzato di elementi, di natura diversa, che interagiscono in modo coordinato
 - il sistema informativo è di solito suddiviso in sottosistemi (in modo gerarchico o decentrato), più o meno integrati
 - componente (sottosistema) di un'organizzazione, di supporto ad altri sottosistemi
- ◆ "Informativo"
 - tutto è finalizzato alla gestione delle informazioni
 - le interazioni corrispondono principalmente a scambi di informazioni (flussi informativi)
- ◆ Gestione delle informazioni:
 - Raccolta, acquisizione
 - Archiviazione, conservazione
 - Elaborazione, trasformazione, produzione
 - Distribuzione, comunicazione, scambio

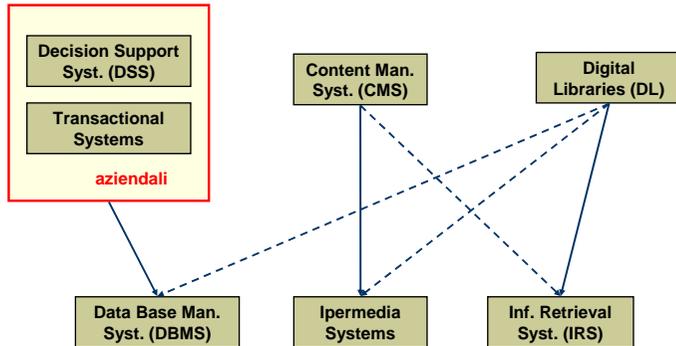
3

Classificazione dei SI

- ◆ Funzione
 - Supporto ai processi aziendali
 - **operativi**: es: gestione di un anagrafe, di un magazzino ...
 - **decisionali**: es: pianificazione delle strategie di marketing, ...
 - Gestione di archivi documentali (DL)
 - es: biblioteca digitale, news, ...
 - Gestione di percorsi informativi (CMS)
 - es. museo virtuale, applicazioni di e-learning
- ◆ Formato dei dati
 - testi, sequenze sonore, video, numeri, ...
- ◆ Grado di strutturazione dei dati gestiti
 - strutturati, semi-strutturati, non strutturati

4

Classificazione (parziale) dei SI



5

L'informazione

Gestione dell'informazione

- ◆ Nelle attività umane, le informazioni vengono rappresentate e scambiate in forme diverse:
 - linguaggio naturale (scritto o parlato, formale o colloquiale, in una lingua o in un'altra)
 - disegni, grafici, schemi
 - numeri e codici
- ◆ Nei sistemi informativi
 - Tendenza a razionalizzare e standardizzare l'organizzazione delle informazioni e delle procedure di gestione
 - ... anche nei sistemi informativi non automatizzati

7

Rappresentazione dell'informazione

- ◆ Nelle attività standardizzate dei sistemi informativi complessi, sono state introdotte col tempo forme di **organizzazione e codifica** delle informazioni
- ◆ Ad esempio, nei servizi anagrafici si è iniziato con registrazioni discorsive e poi
 - nome e cognome
 - estremi anagrafici
 - codice fiscale

8

Informazioni e dati

- ◆ Nei sistemi informatici, le informazioni sono rappresentate in modo essenziale, conciso: attraverso i **dati**
- ◆ (Dal Vocabolario della lingua italiana)
 - **informazione**: notizia, dato o elemento che consente di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere
 - **dato**: ciò che è immediatamente presente alla conoscenza, prima di ogni elaborazione;
(in informatica) elementi di informazione costituiti da simboli che debbono essere elaborati [*interpretati*]

9

Informazioni e dati

- ◆ Perché i dati?
 - La rappresentazione precisa di forme più ricche di informazione e conoscenza è difficile
- ◆ I dati hanno bisogno di essere interpretati
 - Esempio:* `'Mario' '275'`
su un foglio di carta sono due **dati**.
 - Se il foglio di carta viene fornito in risposta alla domanda "A chi mi devo rivolgere per il problema X; qual è il suo numero di telefono?"
allora i dati possono essere interpretati per fornire **informazione** e arricchire la **conoscenza**.

10

Informazione: contenuto + struttura

- ◆ Informazione = Contenuto + Struttura
- ◆ **Contenuto**: l'oggetto che si intende comunicare o rappresentare attraverso:
 - Testo, linguaggio naturale
 - Immagine, linguaggio visuale
 - Suono, linguaggio musicale
- ◆ **Struttura**: l'organizzazione della comunicazione
 - esplicita, attraverso astrazioni, i tratti salienti della rappresentazione
 - fornisce uno schema interpretativo
 - le informazioni sulla struttura sono meta-dati

11

Grado di strutturazione delle informazioni

- ◆ **Informazione strutturata**: molta struttura, contenuto rigidamente codificato
 - Rigida a costruirsi e ad alimentarsi
 - Adatta e flessibile da interrogare per estrarre conoscenza (su proprietà e relazioni delle entità rappresentate)
 - Esempio: i sistemi informativi aziendali
- ◆ **Informazione non strutturata**: molto contenuto (testo), poca struttura
 - Flessibile a costruirsi e ad alimentarsi
 - Difficile da interrogare per estrarre conoscenza
 - Esempio: un romanzo ed il suo indice

12

Struttura e modello dei dati

- ◆ **Modello dei dati:**
 - insieme di meccanismi di astrazione per la rappresentazione di informazioni
- ◆ **Schema e istanze**
 - Lo schema descrive i tipi di proprietà comuni alle varie istanze di una **classe di entità**
 - Ogni istanza rappresenta un'entità mediante un insieme di valori, uno per ogni proprietà dello schema
 - Lo schema impone la **struttura** dell'istanza, i valori delle proprietà ne rappresentano il **contenuto**
- ◆ **Modello dei dati e schemi**
 - la struttura di una classe di informazioni può essere specificata mediante uno schema
 - diversi modelli sono utilizzati per rappresentare schemi (concettuale) e istanze (logico/fisico)

13

Informazione strutturata: modello relazionale

- ◆ **Intuizione**
 - i dati sono organizzati in "tabelle"
 - la tabella è un insieme di "tuple" con attributi di tipi opportuni (numeri, stringhe, date ...)
 - entità e fatti del mondo reale sono rappresentate (in modo molto sintetico) mediante tuple
- ◆ **Tupla (o *ennupla*, o *record*):**
 - insieme di valori, ognuno dei quali corrisponde ad un dato attributo
 - Attributi = struttura
 - Valori = contenuto
- ◆ **Tabella**
 - insieme di tuple
- ◆ **Base di dati**
 - insieme di tabelle

14

Modello relazionale: Esempio

tabella (istanza) tupla attributo valore

Studenti	matricola	cognome	nome	data Nascita
	6554	Rossi	Mario	05/12/1978
	8765	Neri	Paolo	03/11/1976
	9283	Verdi	Luisa	12/11/1979
	3456	Rossi	Maria	01/02/1978

15

Modello relazionale: schemi e istanze

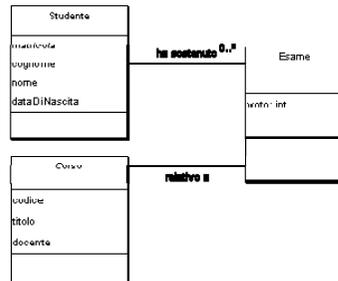
- ◆ **Lo schema:** la descrizione della struttura della tabella
 - stabile nel tempo
 - caratterizza i dati e ne permette l'interpretazione (metadati)
- ◆ **L'istanza:** i valori (cioè i dati) contenuti nella tabella
 - variabile nel tempo

studenti	Matricola	Cognome	Nome	Data di nascita
	6554	Pinco	Pallino	05/12/1978
	8765	Neri	Paolo	03/11/1976
	9283	Verdi	Luisa	12/11/1979
	3456	Rossi	Maria	01/02/1978

16

Informazione strutturata: modello ad oggetti

- ◆ i dati sono organizzati in classi
 - ogni classe rappresenta un insieme di oggetti
 - con un insieme di proprietà
- ◆ le classi sono organizzate in tassonomie
- ◆ oggetti diversi sono correlati sulla base di identificatori



17

Informazione non strutturata: documenti

In an attempt to provide solutions to address these issues, a pilot program is underway in the Environmental Protection Agency (EPA) to make use of semantic technologies to connect information from the Center for Disease Control and Prevention (CDC) and the EPA, as well as from their state partners, in ways that can move the EPA further down the path to answering the public's question: "Is my child safe from environmental toxins?" (Sternag, 2002) While the focus of this pilot is primarily technical in nature, the successful deployment of more expressive capabilities holds enormous human considerations, offering great potential for improving the health and wellbeing of millions of children across the country. Quickly identifying potential toxic exposures, knowing the location and severity of affected sites, and effectively prioritizing environmental cleanup are just three of the most basic priorities for agencies and industry and for the beneficiaries of these efforts – children, their parents, and all other members of society.

This story is one illustration of the tremendous IT challenges that the federal government faces. The complexity of the federal government, the size of its data stores, and its interconnected nature to state, local, and tribal government agencies as well as, increasingly, to private enterprise and Non-governmental Organizations (NGOs) has placed increasing pressure on finding faster, cheaper, and more reliable methods of connecting systems, applications, and data. Connecting these islands of information within and between government agencies and third parties is seen as a key step to improving government services, streamlining processes and logistics, increasing the reliable operation of complex machinery, advancing people's health and welfare, enabling more effective defense capabilities, and ensuring the safety of our nation.

Widespread information interoperability is one of the benefits that many researchers, thought-leaders, and practitioners see for semantic technologies. But by no means is it the only benefit. Building on top of the notion of other more accessible and automatic information, for greater capabilities such as intelligent search, intelligent reasoning, and truly adaptive computing are seen as coming ever closer to reality.

Although semantic is the field of semantic computing has been at work for years, the approval of two new protocols by the World Wide Web Consortium (W3C) early in 2004 marked an important milestone in the commercialization of semantic technologies, also spurring development toward the goal of the Semantic Web. In the words of the W3C, "The goal of the Semantic Web initiative is as broad as that of the Web: to create a universal medium for the exchange of data." "The Semantic Web is a vision: the idea of having data on the web defined and linked in ways so that it can be used by machines – not just for display purposes – but for automation, integration and reuse of data across various applications, and thus fully harness the power of information semantics."

These new capabilities in information technology will not come without significant work and investment by early pioneers. Elements central to this revolution include distributed relational databases or moving from procedural programming techniques to object-oriented approaches. It will take a lot of time for people to understand the nuances and architecture of semantic-based approaches. But as people grasp the full power of these new technologies and approaches, a full generation of innovators will produce impressive results for a number of existing IT problem areas. Successive innovators will ultimately lead to dramatic new capabilities that fundamentally change the way we share and exchange information across users, systems, and

"The World Wide Web Consortium's Semantic Web Activity Statement: <http://www.w3.org/2004/03/sws/>"

Modelli dei dati con poca strutturazione:
 - sequenza di caratteri
 - sequenza di parole
 - al più raggruppate in sezioni, capitoli, ...

1. The Semantic Web is not some magical piece of artificial intelligence. The concept of machine-understandable documents does not imply some form of magical artificial intelligence that allows machines to comprehend human mumbings. It only indicates a machine's ability to solve a well-defined problem by performing well-defined operations on existing well-defined data (Benners-Lee, Handler, and Lassila, 2001). Current search engines perform capabilities that would have been magical 20 years ago, but that we recognize now as being the result of IP protocols, HTML, the concept of websites, web pages, links, graphical browsers, innovative search and ranking algorithms, and, of course, a large number of incredibly fast servers and equally large and fast disk storage arrays. Semantic Web capabilities will likewise be the result of a logical series of interconnected progressions in information technology and knowledge representation formed around a common base of standards and approaches.

2. The Semantic Web is not an existing entity, ready for users to make use of. The Semantic Web currently exists as a vision, albeit a promising and captivating one. Similar to the current Web, the Semantic Web will be formed through a combination of open standard and proprietary protocols, frameworks, technologies, and services. The W3C-approved standards – XML, RDF, and OWL – form the base protocols. New data schemas and contract mechanisms, built using these new protocols, will arise around communities of interest, industry, and intent; some will be designed carefully by experienced data architects and formally recognized by established standards bodies; others will appear from out of nowhere and gain widespread acceptance overnight. A host of new technologies and services will appear such as semantically-aware content publishing tools, context modeling tools, mediation, inference, and routing engines; data-cleaning and thesaurus services; and new authentication and verification components. Although various elements of the vision already exist, rollout of these technologies, coordination amidst competitive forces, and fulfillment of the vision will take many years.

18

Informazione semi-strutturata

- ◆ Vari compromessi possibili fra contenuto e struttura
 - rappresentazioni dell'informazione che cercano di combinare i vantaggi dei due estremi
- ◆ I linguaggi di markup permettono di
 - introdurre struttura nel documento, ma in modo flessibile
 - annotare i contenuti dei documenti
- ◆ Esempi
 - HTML: descrizione delle modalità di presentazione degli elementi del documento
 - XML: caratterizzazione semantica degli elementi del documento

19

Dati semi-strutturati in XML

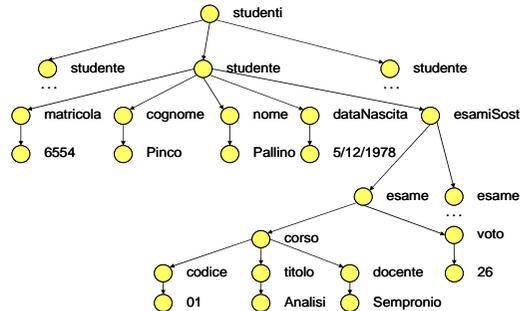
```
<xml version="1.0" ?>
<studenti>
  <studente>
    <matricola>6554</matricola>
    <cognome>Pinco</cognome>
    <nome>Pallino</nome>
    <dataDiNascita>15/12/1978</dataDiNascita>
    <esamiSostenuti>
      <esame>
        <corso>
          <codice>01</codice>
          <titolo>Analisi</titolo>
          <docente>Giacomo</docente>
          <voto>26</voto>
        </corso>
      </esame>
    </esamiSostenuti>
  </studente>
  <studente>
    <matricola>6500</matricola>
    ...
  </studente>
  ...
</studenti>
```

Il documento contiene informazioni sulla sua struttura: alcune parti sono associate a proprietà/attributi (di cui rappresentano il valore)

20

Struttura del documento: il modello di XML

- ◆ Il modello sottostante è un albero di elementi
 - i dati sono organizzati in strutture gerarchiche (alberi)
 - ogni albero ha un insieme di nodi (elementi)
 - oggetti diversi sono correlati sulla base di relazioni di contenimento



21

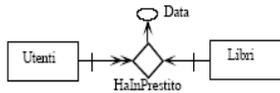
Dati semi-strutturati e Web

- ◆ Il Web come SI globale ed **eterogeneo**:
 - Una quantità enorme di dati, che codificano informazioni eterogenee: disparati argomenti e domini informativi
 - Le informazioni sono create e usate da un numero altissimo di persone, per vari scopi e con differenti criteri di modellazione
- ◆ Quale grado di strutturazione? Quale modello?
 - Approccio strutturato è impraticabile: quale schema?
 - Approccio non strutturato non permette una reale condivisione/compressione delle informazioni
- ◆ Vantaggi dell'approccio semi-strutturato
 - permette di introdurre struttura in modo flessibile: documenti auto-descrittivi
 - Strumento per l'interoperabilità semantica se usato in combinazione con ontologie (visione del *Semantic Web*)

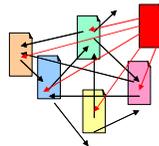
22

Rappresentazione delle associazioni

- ◆ A livello di schema



- ◆ Le relazioni fra istanze si possono rappresentare con riferimenti
 - link fra documenti (modello ipertestuale)
 - proprietà contenenti puntatori ad altri oggetti (modello ad oggetti)



23

Associazioni fra tuple nel modello relazionale

Studenti	matricola	cognome	nome	dataNascita
	6554	Rossi	Mario	05/12/1978
	8765	Neri	Paolo	03/11/1976
	9283	Verdi	Luisa	12/11/1979
	3456	Rossi	Maria	01/02/1978

Corsi	codice	titolo	docente
	a01	Analisi	Pinco
	c02	Chimica	Bruni
	c04	Chimica	Verdi

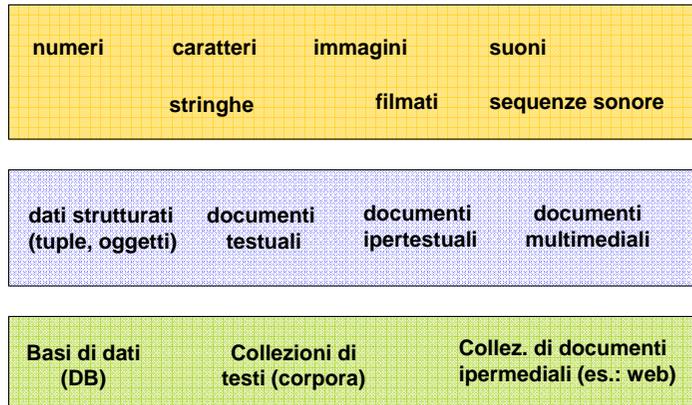
Esami	matricola	voto	lode	corso
	3456	30	1	c04
	3456	24	0	c02
	9283	28	0	a01
	6554	26	0	a01

join

corrispondenza di valori
per attributi logicamente
equivalenti in tabelle
diverse

24

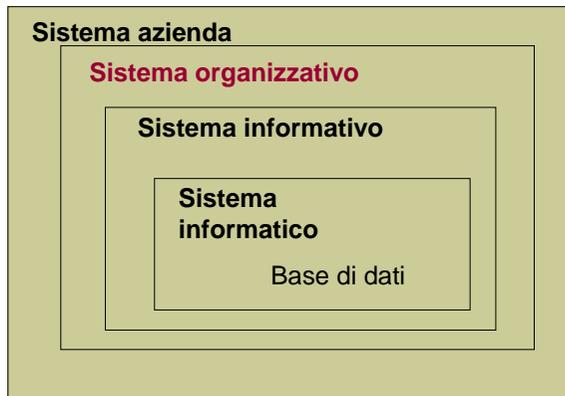
Dati e informazione: livelli di granularità



25

Basi di Dati

La visione dei SI aziendali



27

La visione dei SI aziendali

- ◆ **Sistema organizzativo**
 - l'insieme di risorse e regole per l'uso coordinato di tali risorse, che un'organizzazione impiega per perseguire i propri scopi
- ◆ **Sistema Informativo**
 - l'insieme delle informazioni utilizzate, memorizzate, elaborate in una organizzazione per il perseguimento dei propri scopi
- ◆ **Sistema Informatico**
 - un sistema informativo (o una parte di esso) che si avvale di tecnologie informatiche

28

Sistema organizzativo

- ◆ Insieme di risorse e regole per lo svolgimento coordinato di attività (processi) al fine del perseguimento degli scopi dell'azienda
- ◆ Due concetti sono utili nel descrivere un sistema organizzativo:
 - Risorse
 - Processi
- ◆ **Risorsa aziendale**: tutto ciò con cui la azienda opera, sia materiale che immateriale per perseguire i suoi obiettivi
 - persone
 - denaro
 - materiali
 - **informazioni**

29

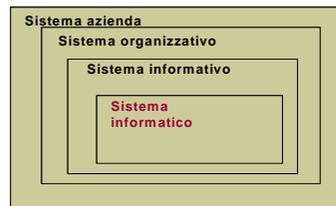
Sistema Informativo

- ◆ Componente (sottosistema) di una organizzazione
 - gestisce (acquisisce, elabora, conserva, produce) le informazioni di interesse (cioè utilizzate per il perseguimento degli scopi dell'organizzazione)
 - ogni organizzazione ha un sistema informativo, eventualmente non esplicitato nella struttura;
- ◆ Architettura:
 - quasi sempre, il SI è di supporto ad altri sottosistemi dell'organizzazione
 - il SI è di solito suddiviso in sottosistemi (in modo gerarchico o decentrato), più o meno fortemente integrati

30

Sistema informatico

- ◆ Porzione automatizzata del sistema informativo:
 - la parte del sistema informativo che gestisce informazioni con tecnologia informatica



- ◆ Il concetto di "sistema informativo" è indipendente da qualsiasi automatizzazione:
 - esistono organizzazioni operanti da secoli la cui ragion d'essere è la gestione di informazioni (p. es. servizi anagrafici e banche)

31

Approccio tipico: Basi di dati

- ◆ Scopo:
 - memorizzare l'informazione necessaria per supportare alcuni processi del sistema organizzativo
- ◆ Strategia di rappresentazione
 - le informazioni sono rappresentate mediante **dati strutturati**, modellati in modo essenziale
- ◆ Modalità di uso
 - Frequenti operazioni di accesso a porzioni di dati ben specificate (singola tupla/record o piccoli insiemi di tuple) durante l'esecuzione dei processi aziendali
 - lettura (interrogazione)
 - scrittura (aggiornamento)
 - Le interrogazioni richiedono risposte rapide ed **esatte**
 - Le operazioni sono spesso effettuate in modo automatico da applicazioni che gestiscono specifici processi aziendali

32

Basi di dati e DBMS

- ◆ Base di dati
 - Accezione generica, metodologica:
 - Insieme organizzato di dati utilizzati per il supporto allo svolgimento delle attività di un ente (azienda, ufficio, persona)
 - Accezione specifica, metodologica e tecnologica:
 - insieme di dati gestito da un DBMS
- ◆ DBMS (DataBase Management System)
 - Sistema software in grado di gestire una base di dati

33

Sistema di gestione di basi di dati (DBMS)

Sistema (**prodotto software**) in grado di gestire **collezioni di dati ...**

- **Grandi**: di dimensioni (molto) maggiori della memoria centrale dei sistemi di calcolo utilizzati
- **Persistenti**: con un periodo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano
- **Condivise**: utilizzate da applicazioni/utenti diversi

... garantendo

- **affidabilità** (resistenza a malfunzionamenti hardware e software)
- **privatizza** (mediante il controllo degli accessi)

Proprietà desiderate:

- **efficienza** (utilizza al meglio le risorse di memoria ed il tempo di uso del sistema)
- **efficacia** (rende produttive le attività dei suoi utilizzatori)

34

Alcuni DBMS in commercio

- ◆ Access
- ◆ SQLServer
- ◆ DB2
- ◆ Oracle
- ◆ Informix
- ◆ Sybase
- ◆ Ingres

35

Schemi e istanze

◆ In ogni base di dati esistono:

- lo **schema**, sostanzialmente invariante nel tempo, che ne descrive la struttura (aspetto intensionale)
 - in una base di dati relazionale, le intestazioni delle tabelle
- l'**istanza**, i valori attuali, che possono cambiare anche molto rapidamente (aspetto estensionale)
 - in una base di dati relazionale, il "corpo" di ciascuna tabella

Ex

36

Architettura dei dati in un DBMS



■ **schema logico**: descrizione della base di dati nel modello logico (es., la struttura della tabella)

■ **schema fisico**: rappresentazione dello schema logico per mezzo di strutture memorizzazione (file)

37

Descrizione dei dati nei DBMS: Modelli dei dati

- ◆ insieme di costrutti utilizzati per organizzare i dati di interesse e descriverne la struttura (e, in parte, la semantica)
- ◆ componente fondamentale: **meccanismi di strutturazione** (o **costruttori di tipo**)
- ◆ il **modello relazionale** prevede il costruttore **relazione**, che permette di definire insiemi di record omogenei

38

Due tipi (principali) di modelli

- ◆ **modelli logici**: utilizzati nei DBMS per l'organizzazione dei dati
 - utilizzati (dai programmi e dagli utenti) per accedere ai dati
 - indipendenti dalle strutture fisiche
 - esempi: **relazionale**, ad oggetti
- ◆ **modelli concettuali**: permettono di rappresentare i tipi di dati (struttura) in modo indipendente da ogni sistema
 - cercano di descrivere i concetti del mondo reale
 - sono utilizzati nelle fasi preliminari di progettazione
 - Esempio più noto: il modello **Entity-Relationship**

39

Il modello relazionale

- ◆ Storia
 - Proposto da Codd nel 1970 per favorire l'indipendenza dei dati
 - Disponibile in DBMS reali nel 1981 (non è facile implementare l'indipendenza con efficienza e affidabilità!)
- ◆ Si basa sul concetto matematico di relazione
 - Le relazioni hanno una naturale rappresentazione in forma di tabelle
- ◆ Permette di specificare sia la rappresentazione dei dati sia le operazioni di accesso ad essi
 - Interrogazioni
 - Modifiche, inserimenti di valori

40

Relazione: tre accezioni

- ◆ **relazione matematica** (*relation*): come nella teoria degli insiemi
- ◆ **relazione** (*relationship*): costituisce un'associazione o correlazione
- ◆ **relazione** secondo il modello relazionale dei dati

41

Relazione matematica, esempio

- $D_1 = \{a, b\}$
- $D_2 = \{x, y, z\}$
- prodotto cartesiano $D_1 \times D_2$

a	x
a	y
a	z
b	x
b	y
b	z

- ◆ una relazione $r \subseteq D_1 \times D_2$

a	x
a	z
b	y

42

Relazione matematica, esempio

$Partite \subseteq string \times string \times int \times int$

Juve	Lazio	3	1
Lazio	Milan	2	0
Juve	Roma	0	2
Roma	Milan	0	1

- ◆ Ciascuno dei domini ha due ruoli diversi, distinguibili attraverso la posizione:
 - La struttura è **posizionale**

43

Struttura non posizionale

A ciascun dominio si associa un nome (**attributo**), che ne descrive il "ruolo"

Casa	Fuori	RetiCasa	RetiFuori
Juve	Lazio	3	1
Lazio	Milan	2	0
Juve	Roma	0	2
Roma	Milan	0	1

44

Tabelle e relazioni

- ◆ Una tabella rappresenta una relazione se
 - i valori di ogni colonna sono fra loro omogenei
 - le righe sono diverse fra loro
 - le intestazioni delle colonne sono diverse tra loro
- ◆ In una tabella che rappresenta una relazione
 - l'ordinamento tra le righe è irrilevante
 - l'ordinamento tra le colonne è irrilevante
- ◆ Il modello è basato sui valori
 - i riferimenti fra dati in relazioni diverse sono rappresentati per mezzo di valori dei domini che compaiono nelle tuple

45

Esempio

- ◆ Base di dati universitaria
- ◆ Studenti
 - nome, cognome, matricola, data di nascita
- ◆ Corsi
 - codice, nome del corso, nome del docente
- ◆ Esami sostenuti
 - corso, studente, voto

46

Esempio

- ◆ Studenti
 - nome: stringa
 - cognome: stringa
 - matricola: intero
 - data di nascita: data
- ◆ Corsi
 - codice: stringa
 - nome: stringa
 - nome del docente: stringa
- ◆ Esami sostenuti
 - corso: "riferimento" ad un corso
 - studente: "riferimento" ad uno studente
 - voto: intero
 - lode: si/no

47

Base di Dati

Studenti	matricola	cognome	nome	dataNascita
	6554	Rossi	Mario	05/12/1978
	8765	Neri	Paolo	03/11/1976
	9283	Verdi	Luisa	12/11/1979
	3456	Rossi	Maria	01/02/1978

Corsi	codice	titolo	docente
	a01	Analisi	Pinco
	c02	Chimica	Bruni
	c04	Chimica	Verdi

Esami	matricola	voto	lode	corso
	3456	30	1	c04
	3456	24	0	c02
	9283	28	0	a01
	6554	26	0	a01

48

Vincoli di Integrità

- ◆ Regole imposte sui valori della base di dati
- ◆ Vincoli di **chiave** (su una sola tabella)
 - chiave: identificatore per le tuple
 - es: "matricola" è una chiave per "Studenti"
- ◆ Vincoli di **tupla** (su una sola tabella)
 - predicati sui valori delle ennuple
 - es: (voto>=18 and voto<=30)
- ◆ Vincoli di **riferimento** (fra tabelle)
 - assenza di riferimenti inesistenti
 - es: esistono esami solo per gli studenti della bd

49

Chiavi

- ◆ Esiste sempre una chiave (almeno implicita)
 - Ogni relazione ha come superchiave l'insieme degli attributi su cui è definita: non può contenere tuple distinte ma con valori uguali
- ◆ **Importanza**
 - l'esistenza delle chiavi garantisce l'accessibilità a ciascun dato della base di dati
 - le chiavi permettono di correlare i dati in relazioni diverse:
 - il modello relazionale è basato su valori

50

Una Base di Dati Scorretta

Studenti	matricola	cognome	nome
	6554	Rossi	Mario
	78787	Neri	Piero
	78787	Bianchi	Luca

unicità della matricola

Esami	matricola	voto	lode	corso
	6554	32	0	a01
	78787	30	1	c02
	6554	27	1	d03
	1122	24	0	c04

voti scorretti

riferimento scorretto

51

Linguaggi di interrogazione (Query Languages)

52

Linguaggio di interrogazione

- ◆ Per richiedere al DBMS il recupero di dati
 - eventualmente derivati tramite composizione e aggregazione
- ◆ Ogni modello ha i suoi linguaggi
 - DBMS relazionali: **SQL, QBE**
 - DBMS a oggetti: **OQL**
 - XML: **XPath, XQuery**

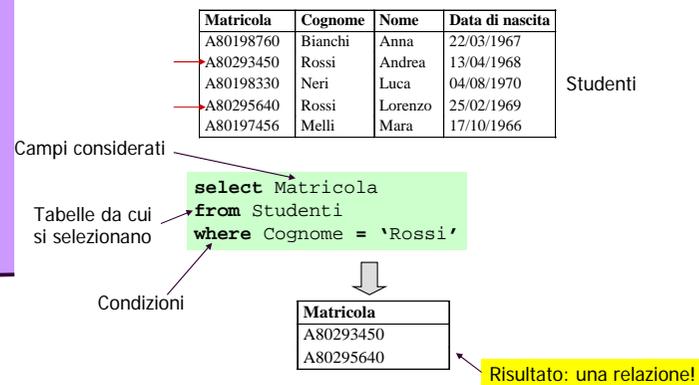
53

Un linguaggio per DB relazionali: Structured Query Language (SQL)

- ◆ Linguaggio dichiarativo
 - Le interrogazioni esprimono **cosa** si cerca e **non come**
 - Il **come** è prodotto dall'interprete SQL del DBMS che traduce l'interrogazione SQL nelle operazioni interne sui dati che producono la risposta
 - L'utente non deve quindi preoccuparsene
- ◆ Linguaggio orientato agli insiemi (tabelle)
 - Input = insieme di tabelle
 - Output = tabella (insieme di tuple)
- ◆ Le operazioni di interrogazione vengono specificate con l'istruzione **select**

54

Interrogazioni in SQL: l'istruzione SELECT



55

DB di esempio

Maternità	Madre	Figlio	Persone		
			Nome	Età	Reddito
	Luisa	Maria	Andrea	27	21
	Luisa	Luigi	Aldo	25	15
	Anna	Olga	Maria	55	42
	Anna	Filippo	Anna	50	35
	Maria	Andrea	Filippo	26	30
	Maria	Aldo	Luigi	50	40
Paternità	Padre	Figlio			
	Sergio	Franco	Franco	60	20
	Luigi	Olga	Olga	30	41
	Luigi	Filippo	Sergio	85	35
	Franco	Andrea	Luisa	75	87
	Franco	Aldo			

56

SELECT: selezione (da una tabella)

QUERY: *persone con meno di 30 anni*

```
SELECT nome, eta, reddito
FROM persone
WHERE eta < 30
```

Persone		
Nome	Età	Reddito
Andrea	27	21
Aldo	25	15
Maria	55	42
Anna	50	35
Filippo	26	30
Luigi	50	40
Franco	60	20
Olga	30	41
Sergio	85	35
Luisa	75	87

Nome	Età	Reddito
Andrea	27	21
Aldo	25	15
Filippo	26	30

57

SELECT: proiezione (da una tabella)

QUERY: *Nome e età di tutte le persone*

```
SELECT nome, eta
FROM persone
```

Persone		
Nome	Età	Reddito
Andrea	27	21
Aldo	25	15
Maria	55	42
Anna	50	35
Filippo	26	30
Luigi	50	40
Franco	60	20
Olga	30	41
Sergio	85	35
Luisa	75	87

Nome	Età
Andrea	27
Aldo	25
Maria	55
Anna	50
Filippo	26
Luigi	50
Franco	60
Olga	30
Sergio	85
Luisa	75

58

Selezione e proiezione (da una tabella)

QUERY: *Nome e reddito delle persone con meno di 30 anni*

```
SELECT nome, reddito
FROM persone
WHERE eta < 30
```

Persone		
Nome	Età	Reddito
Andrea	27	21
Aldo	25	15
Maria	55	42
Anna	50	35
Filippo	26	30
Luigi	50	40
Franco	60	20
Olga	30	41
Sergio	85	35
Luisa	75	87

Persone	
Nome	Reddito
Andrea	21
Aldo	15
Filippo	30

59

Interrogazione non testuale: Query By Example (QBE)

Microsoft Access

File Modifica Visualizza Inserisci Query Strumenti Finestra 2

CorsiAPianoTerra: Query di selezione

Campo:	Corso	Aula	Piano	
Tabella:	Corsi	Corsi	Aule	
Ordinamento:				
Mostra:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Criteri:			"Terra"	
Oppure:				

Pronto NUM

60

"Fusione orizzontale": Join

Permette di integrare contenuti di tabelle diverse

Es: Nome, età e padre per tutte le persone

```
SELECT X.nome, X.eta, P.padre
from persone X, paternita P
where P.figlio = x.nome
```

Maternità	Madre	Figlio	Persone		
			Nome	Età	Reddito
	Luisa	Maria	Andrea	27	21
	Luisa	Luigi	Aldo	25	15
	Anna	Olga	Maria	55	42
	Anna	Filippo	Anna	50	35
	Maria	Andrea	Filippo	26	30
	Maria	Aldo	Luigi	50	40
Paternità	Padre	Figlio			
	Sergio	Franco	Franco	60	20
	Luigi	Olga	Olga	30	41
	Luigi	Filippo	Sergio	85	35
	Franco	Andrea	Luisa	75	87
	Franco	Aldo			

61

Join

Nome, padre e madre per tutte le persone

```
SELECT P.padre, M.madre
FROM paternita P, maternita M
WHERE P.figlio = M.figlio
```

```
SELECT P.padre, M.madre
FROM paternita P
JOIN maternita M
ON P.figlio= M.figlio
```

Maternità	Madre	Figlio	Persone		
			Nome	Età	Reddito
	Luisa	Maria	Andrea	27	21
	Luisa	Luigi	Aldo	25	15
	Anna	Olga	Maria	55	42
	Anna	Filippo	Anna	50	35
	Maria	Andrea	Filippo	26	30
	Maria	Aldo	Luigi	50	40
Paternità	Padre	Figlio			
	Sergio	Franco	Franco	60	20
	Luigi	Olga	Olga	30	41
	Luigi	Filippo	Sergio	85	35
	Franco	Andrea	Luisa	75	87
	Franco	Aldo			

62

Selezione, proiezione e join

I padri di persone che guadagnano più di venti milioni

```
select distinct padre
from persone, paternita
where figlio = nome and reddito > 20
```

Maternità	Madre	Figlio	Persone		
			Nome	Età	Reddito
	Luisa	Maria	Andrea	27	21
	Luisa	Luigi	Aldo	25	15
	Anna	Olga	Maria	55	42
	Anna	Filippo	Anna	50	35
	Maria	Andrea	Filippo	26	30
	Maria	Aldo	Luigi	50	40
Paternità	Padre	Figlio			
	Sergio	Franco	Franco	60	20
	Luigi	Olga	Olga	30	41
	Luigi	Filippo	Sergio	85	35
	Franco	Andrea	Luisa	75	87
	Franco	Aldo			

63

Ordinamento del risultato

Nome e reddito delle persone con meno di trenta anni in ordine alfabetico

```
select nome, reddito
from persone
where eta < 30
order by nome
```

Persone		→	Persone	
Nome	Reddito		Nome	Reddito
Andrea	21		Aldo	15
Aldo	15		Andrea	21
Filippo	30		Filippo	30

64

DISTINCT

“Cognomi degli studenti, in ordine alfabetico”

```
SELECT DISTINCT cognome
FROM Studenti
ORDER BY cognome;
```

65

Unione

“Cognomi delle persone”

```
SELECT cognome AS cognomePersona
FROM Professori

UNION

SELECT cognome
FROM Studenti;
```

66

Differenza

“Cognome e nome dei professori ordinari che non supervisionano tesi triennali”

```
SELECT cognome AS cognomeProf, nome AS nomeProf
FROM Professori
WHERE qualifica='ordinario'
```

EXCEPT

```
SELECT cognome, nome
FROM Studenti JOIN Professori ON relatore=cod
WHERE ciclo='laurea tr.';
```

67

Differenza: alternativa 1

- ◆ Alcune versioni di SQL non supportano direttamente l'operazione di differenza:
 - bisogna riformulare la query utilizzando il concetto di **negazione**:

“Professore ordinario che non supervisiona tesi triennali”
= “Professore ordinario che **NON APPARTIENE** all'insieme dei professori che supervisionano tesi triennali”

```
SELECT Professori.cognome, Professori.nome
FROM Professori
WHERE Professori.qualifica='ordinario'
AND Professori.cod NOT IN
  (SELECT Studenti.relatore
   FROM Studenti
   AND Studenti.ciclo='laurea tr.');
```

68

Differenza: alternativa 2

“Professore ordinario che non supervisiona tesi triennali”
= “Professore ordinario tale che **NON ESISTE** alcuna tesi triennale da egli supervisionata”

```
SELECT Professori.cognome, Professori.nome
FROM Professori
WHERE Professori.qualifica='ordinario' AND NOT EXISTS
  (SELECT *
   FROM Studenti WHERE Studenti.relatore=Professori.cod
   AND Studenti.ciclo='laurea tr.');
```

69

“Fusione verticale”: operatori di aggregazione

- ◆ Permettono di costruire interrogazioni che coinvolgono più tuple insieme
- ◆ Ad esempio:
 - contare il numero di studenti che hanno superato un esame, trovare il massimo voto preso da uno studente ad un esame, calcolare la media di uno studente
- ◆ **operatori aggregati:**

count	conteggio righe
sum	somma
max	massimo
min	minimo
avg	media

70

Operatori aggregati: COUNT

- Il numero di figli di Franco

```
select count(*) as NumFigliDiFranco
from Paternita
where Padre = 'Franco'
```

- ◆ l'operatore aggregato (**count**) viene applicato al risultato dell'interrogazione:

```
select *
from Paternita
where Padre = 'Franco'
```

Paternità	Padre	Figlio
	Sergio	Franco
	Luigi	Olga
	Luigi	Filippo
	Franco	Andrea
	Franco	Aldo

NumFigliDiFranco
2

71

Conteggio, massimo, media...

Tabella esami

Studente	Voto	Corso
A80198760	28	01
A80293450	30	04
A80198760	27	03
A80293450	25	03
A80293450	21	05

```
select count(*) from esami
where Studente = 'A80293450'
```

```
select max(Voto) from esami
where Studente = 'A80293450'
```

```
select avg(Voto) from esami
where Studente = 'A80293450'
```

Informazioni sui voti dello studente
A80293450

```
select count(distinct Studente)
from esami
```

Conteggio di quanti studenti hanno
fatto almeno un esame

72

Operatori aggregati e gruppi

Le funzioni possono essere applicate a gruppi

GROUP BY listaAttributi

Es.: Il numero di figli di ciascun padre

```
select padre, count(*) AS NumFigli  
from paternita  
group by Padre
```

paternita

Padre	Figlio
Sergio	Franco
Luigi	Olga
Luigi	Filippo
Franco	Andrea
Franco	Aldo

Padre	NumFigli
Sergio	1
Luigi	2
Franco	2

73