

Informatica **U**manistica

Basi di dati

Dino Pedreschi

Dipartimento di Informatica

a.a. 2011–2012



UNIVERSITÀ DI PISA

Ci sono dei fondamenti dell'informatica che dovrebbero stare nel bagaglio dell'umanista?

- ◆ Si danno tante risposte a questa domanda, molte di tipo tecnologico-pratico
- ◆ La risposta che questo corso cerca di dare è metodologica
- ◆ La nostra convinzione è che un(a) umanista dovrebbe padroneggiare i principi per la rappresentazione dell'informazione e della conoscenza

La tecnologia evolve rapidamente, ma ..

- ◆ **La tecnologia avanzata di oggi sarà la tecnologia superata di domani**
- ◆ **Per comprendere (e magari generare) i cambiamenti tecnologici occorre una profonda conoscenza metodologica e scientifica di base**

- ◆ ***Quelli che s'innamora di pratica senza scienza son come 'I nocchieri ch'entra in navilio senza timone o bussola, che mai ha certezza dove si vada.***
- ◆ **Leonardo da Vinci, *Frammenti letterari e filosofici*, Barbera, Firenze, 1809**

Obiettivi del corso

- ◆ Fornire le conoscenze e le competenze necessarie alla **rappresentazione** e al **trattamento** consapevole delle **informazioni** pertinenti alle scienze umane.
- ◆ Il corso prevede un mix di fondamenti di portata generale e di esercitazioni pratiche relative all'uso di tecnologie informatiche per la rappresentazione dell'informazione.

Quale informazione?

- ◆ **Informazione = Contenuto + Struttura**
- ◆ **Contenuto:** l'oggetto che si intende comunicare/rappresentare attraverso
 - Testo, linguaggio naturale
 - Immagine, linguaggio visuale
 - Suono, linguaggio audiovisivo
- ◆ **Struttura:** la organizzazione della comunicazione, che esplicita, attraverso **astrazioni**, i tratti salienti della rappresentazione

Informazione = contenuto + struttura

◆ Informazione non strutturata

- Molto contenuto, poca struttura
- Un romanzo e il suo indice (magari solo capitoli numerati)

◆ Informazione strutturata

- Molta struttura, contenuto rigidamente codificato
- I sistemi informativi delle organizzazioni, le basi di dati

Informazione strutturata vs. non strutturata

◆ Due situazioni estreme

◆ Informazione strutturata

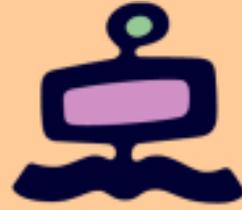
- - Rigida a costruirsi e ad alimentarsi
- + Adatta e flessibile ad interrogarsi per estrarre conoscenza

◆ Informazione non strutturata (testo)

- + Flessibile a costruirsi e ad alimentarsi
- - Difficile ad interrogarsi per estrarre conoscenza

Fra i due estremi ...

- ◆ **L'informazione semi-strutturata**, vari possibili compromessi fra contenuto e struttura
- ◆ **Rappresentazioni dell'informazione**, dei documenti multimediali, che cercano di combinare i vantaggi dei due estremi.
- ◆ **Sullo sfondo, il Web**, una sorgente sterminata di informazione semi-strutturata (a struttura crescente) da cui è difficile estrarre conoscenza



Informatica**U**manistica

Una rapida panoramica sul corso

*Dall'informazione strutturata alla
informazione semi-strutturata*



UNIVERSITÀ DI PISA

Modello dei Dati

- ◆ Insieme di **meccanismi di astrazione** per la rappresentazione di informazioni

Modello relazionale dei dati

- ◆ i dati sono organizzati in “tabelle”
- ◆ la tabella è un insieme di “record” (ennuple)
- ◆ con un insieme di attributi
- ◆ di tipi opportuni (numeri, stringhe, date ...)
- ◆ i dati in tabelle diverse sono correlati sulla base dei valori

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

esami

Studente	Voto	Corso
3456	30	04
3456	24	02
9283	28	01
6554	26	01

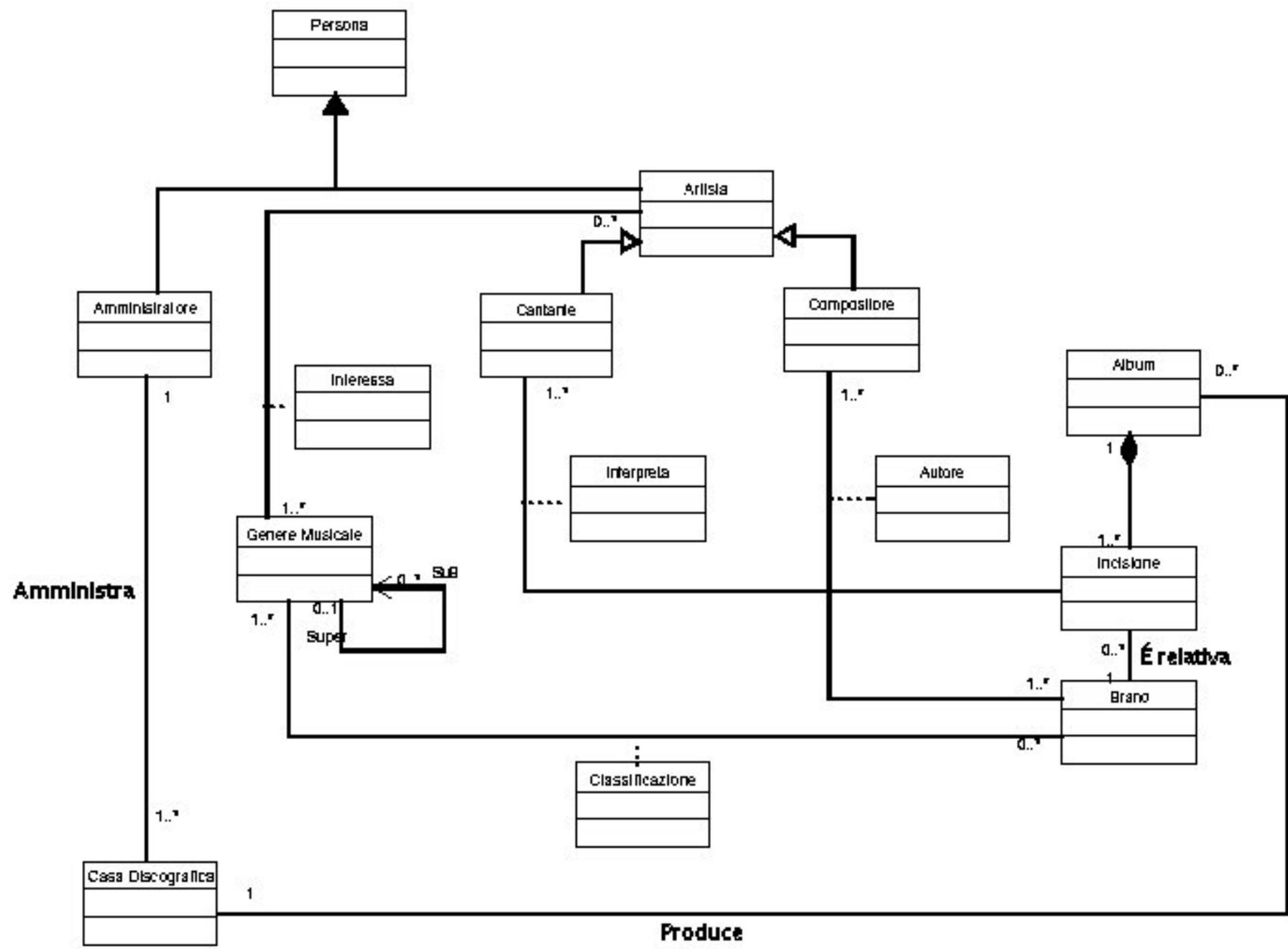
corsi

Codice	Titolo	Docente
01	Analisi	Sempronio
02	Chimica	Bruni
04	Chimica	Verdi

Modello concettuale dei dati

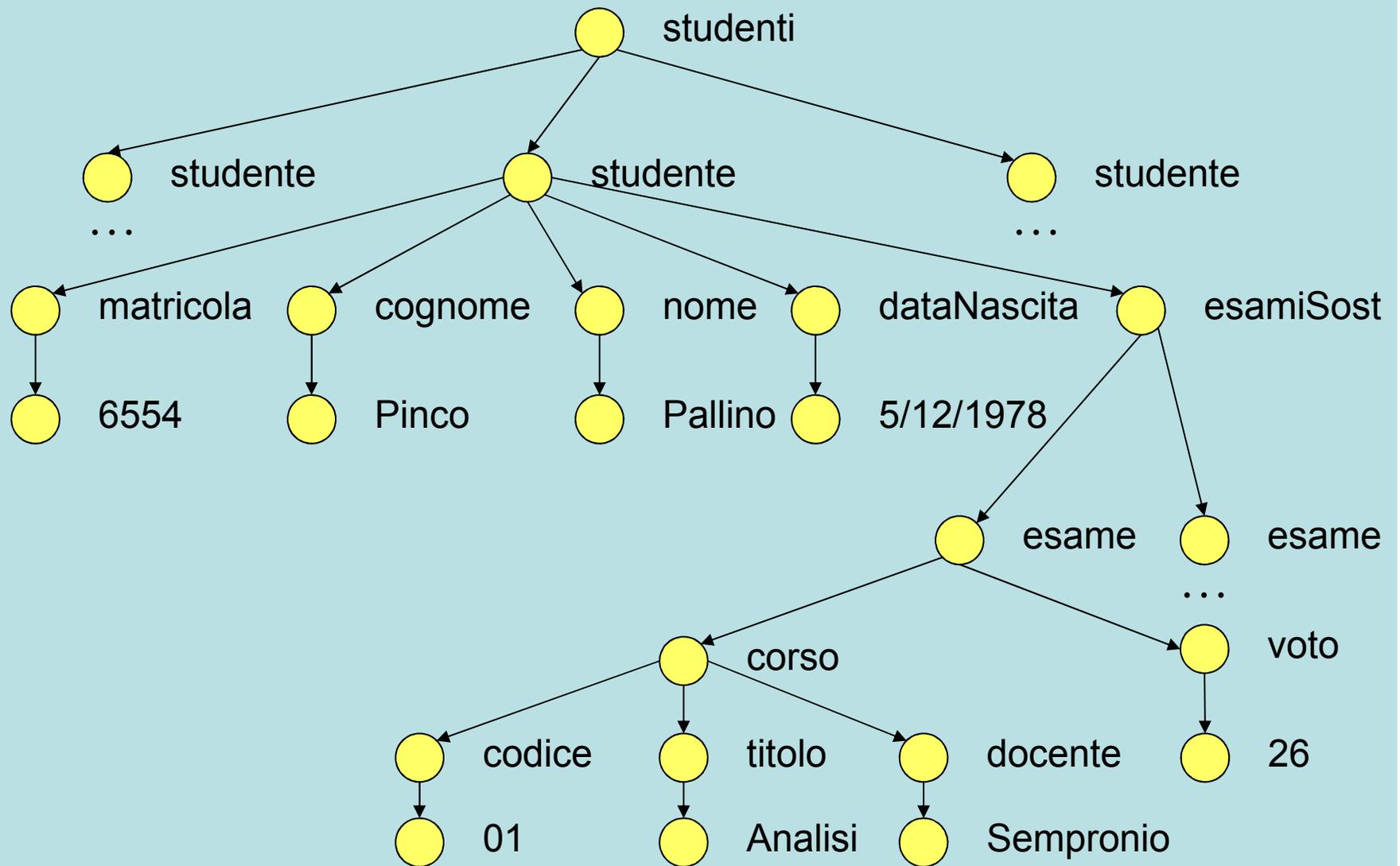
- ◆ i dati sono organizzati in classi
- ◆ ogni classe rappresenta un insieme di oggetti
- ◆ con un insieme di proprietà
- ◆ le classi sono organizzate in ontologie
- ◆ oggetti diversi sono correlati sulla base di identificatori





Modello semi-strutturato: XML

- ◆ **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**



```
<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>
</studenti>
```

In ogni modello esistono

- ◆ **Lo schema:** la descrizione della struttura

- stabile nel tempo

- ◆ **L'istanza:** i valori (cioè i dati)

- 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

Linguaggio di interrogazione

- ◆ Linguaggio per richiedere al DBMS (Database Management System) il recupero di informazione
- ◆ Ogni modello ha i suoi linguaggi
- ◆ DBMS relazionali: **SQL, QBE**
- ◆ DBMS a oggetti: **OQL**
- ◆ XML: **XPath, XQuery**

Schema e Istanza

◆ Che cosa del modello deve conoscere lo sviluppatore dell'applicazione ?

- il modello logico (e non il modello fisico)
- la struttura dei dati e non necessariamente i dati veri e propri

◆ In altri termini

- lo "schema"
- e non necessariamente l'"istanza"

In ogni modello esistono

- ◆ **Lo schema:** la descrizione della struttura

- stabile nel tempo

- ◆ **L'istanza:** i valori (cioè i dati)

- 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

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

esami

Studente	Voto	Corso
3456	30	04
3456	24	02
9283	28	01
6554	26	01

corsi

Codice	Titolo	Docente
01	Analisi	Sempronio
02	Chimica	Bruni
04	Chimica	Verdi

SQL: Un esempio

SELECT Matricola, Cognome, AVG(Voto)

FROM Studenti, Esami

WHERE Matricola = Studente

GROUP BY Matricola, Cognome

Matricola	Cognome	AVG(voto)
6554	Pinco	26
8765	Neri	28
3456	Rossi	27

Sommario

◆ Introduzione

- Sistema Informativo e Sistema Informatico

◆ DBMS

◆ Modello logico e modello fisico

- Panoramica sui modelli
- Schema e Istanza
- Linguaggi per basi di dati

◆ Indipendenza dei dati