

Informatica per le scienze umane

Mirco Nanni
ISTI-CNR, Pisa

CdL in Lettere – A.A. 2008/2009

Docente

Mirco Nanni

- mirco.nanni@isti.cnr.it
- KDD Laboratory, ISTI - CNR, Pisa
- Tel.: 050-3152348
- <http://ercolino.isti.cnr.it/mirco>

Ricevimento

- Giovedì ore 10.00-12.00
- ISTI-CNR, Area della Ricerca CNR di Pisa
(ingresso 20, 1° piano, stanza C45)

Co-docente di laboratorio

◆ Mirko Tavosanis

- Ricercatore presso il Dipartimento di Studi Italianistici

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**

I metodi dell'informatica per ...

- ◆ **descrivere dati, informazione, conoscenza**
- ◆ **accedere a questi dati, informazione, conoscenza conservati:**
 - nei sistemi informativi
 - nelle basi di dati
 - sul Web
 - nelle biblioteche digitali
 - ...
 - nel patrimonio informativo digitale che cresce intorno a noi

La tecnologia evolve rapidamente, ma ...

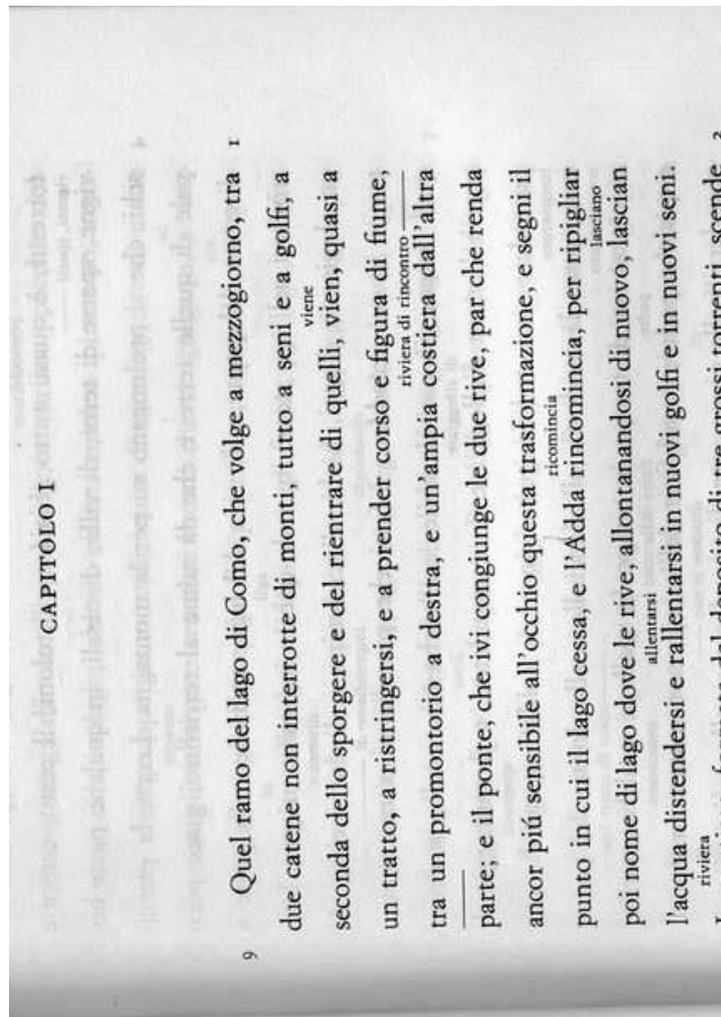
- ◆ **La tecnologia avanzata di oggi sarà la tecnologia superata di domani**
- ◆ **Per comprendere (e magari generare) i cambiamenti tecnologici occorre una profonda conoscenza metodologica e scientifica di base**

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**



Scheda bibliografica

Manzoni, Alessandro.

I promessi sposi / Alessandro Manzoni ; a cura di Natalino Sapegno. - Firenze : Le Monnier, 2005. - xvi, 728 p. ; 24 cm.

ISBN 978-88-00-41298-8

Sapegno, Natalino

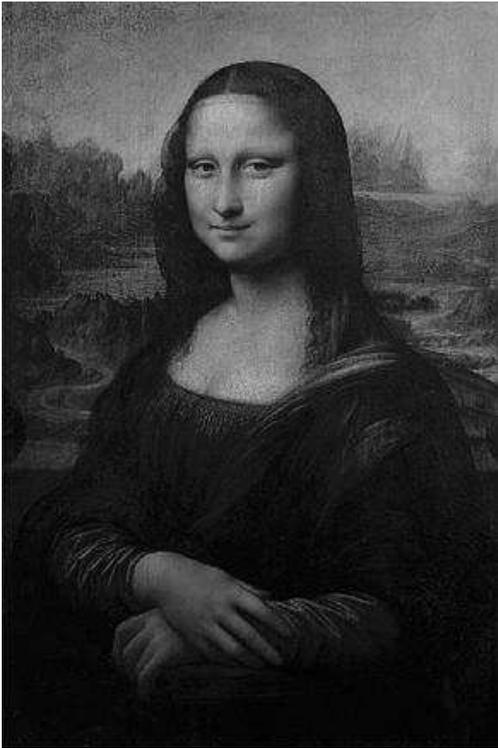
Spartito musicale



Spartito musicale ?



Contenuti diversi, strutture simili



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

- Ognuna con lati positivi (+) e negativi (-)

◆ 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

Contenuti (e struttura ...) del corso

Informazione strutturata

- ◆ **Le basi di dati relazionali**
- ◆ **L'algebra relazionale e il linguaggio di interrogazione SQL (*Structured Query Language* - cenni)**
- ◆ **La progettazione concettuale dei dati e le ontologie**
- ◆ **La sperimentazione con un DBMS**

Informazione semi-strutturata

- ◆ **La descrizione logica dei documenti nel modello semi-strutturato**
- ◆ **I linguaggi di marcatura (*mark-up*)**
- ◆ **Il modello dei dati XML (*eXtended Mark-up Language*)**
- ◆ **I linguaggi di interrogazione per XML**

Informazione non strutturata

- ◆ **Il linguaggio HTML (*Hypertext Mark-up Language*) per la presentazione dei documenti**
- ◆ **Informazione testuale e information retrieval (cenni)**
- ◆ **I documenti Web e i motori di ricerca (cenni)**

Materiale didattico

◆ Lucidi: in collaborazione con

- Domenico Saccà, Università della Calabria
- Gianni Mecca, Università della Basilicata
- Antonio Albano e Paolo Manghi, Università di Pisa
- Dino Pedreschi, Università di Pisa

◆ Dispense ed esercizi

◆ Materiale disponibile on-line

- <http://www.cli.di.unipi.it/doku>
➔ Link: “Informatica per le scienze umane”

Sito web del corso

www.cli.di.unipi.it/doku/doku.php/informaticaperlescienzeumane/start

The screenshot shows a Firefox browser window displaying the DidaWiki page for the course 'Informatica per le scienze umane - A.A. 2007-2008'. The page title is 'Informatica per le scienze umane - A.A. 2007-2008 [DidaWiki] - Firefox'. The address bar shows the URL 'http://www.cli.di.unipi.it/doku/doku.php/informaticaperlescienzeumane/start'. The page content includes a navigation menu with 'Modifica questa pagina' and 'Revisioni precedenti'. The main content area is titled 'Informatica per le scienze umane - A.A. 2007-2008' and contains sections for 'Informazioni generali', 'Docenti', and 'Assistenti'. The 'Informazioni generali' section lists course details such as 'Corso di Laurea in Lettere, Facoltà di Lettere e Filosofia' and 'Settore scientifico-disciplinare: INF/01'. The 'Docenti' section lists 'Mirco Nanni' with contact information and 'Assistenti' section lists 'Alessandro Lenci' and 'Mirko Tavosanis'. A sidebar on the right contains an 'Indice' (Index) with links to various page sections like 'Informazioni generali', 'Docenti', 'Orario e aula', 'Calendario delle lezioni', 'Materiale didattico', 'Testi d'esame', 'Lucidi del docente', 'Esercizi e verifiche precedenti', 'Modalità d'esame', 'Appelli e verifiche', and 'FAQ - Frequently Asked Questions'. The page also features a search bar and a 'Cerca' button.

Testi di consultazione

- ◆ Atzeni, Ceri, Paraboschi, Torlone. *Basi di dati*. McGraw-Hill, 2002.
- ◆ Albano. *Sistemi per l'archiviazione e il recupero delle informazioni*. dispensa, 1999.

Una rapida panoramica sul corso

*Dall'informazione strutturata alla
informazione semi-strutturata*

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**

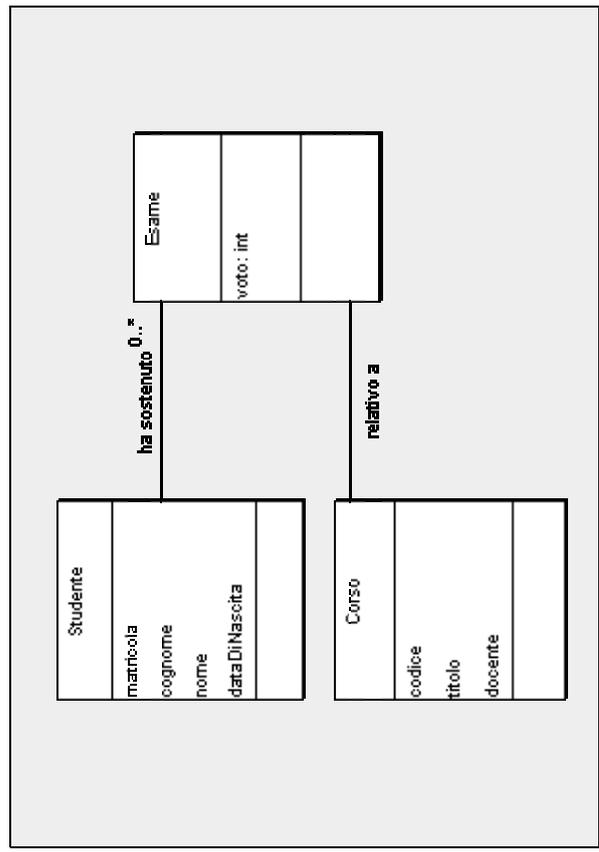
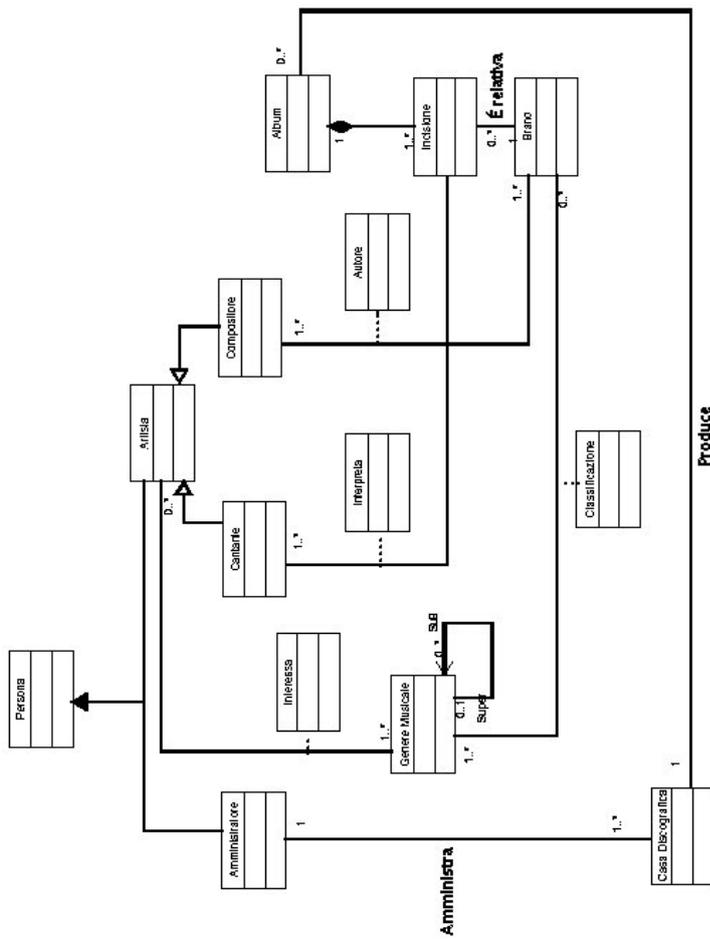
studenti	Matricola	Cognome	Nome	Data di nascita
	6554	Pinco	Pallino	05/12/1978
	8765	Neri	Paolo	03/11/1976
	9283	Verdi	Luisa	12/11/1979
	3456	Rossi	Maria	01/02/1978

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

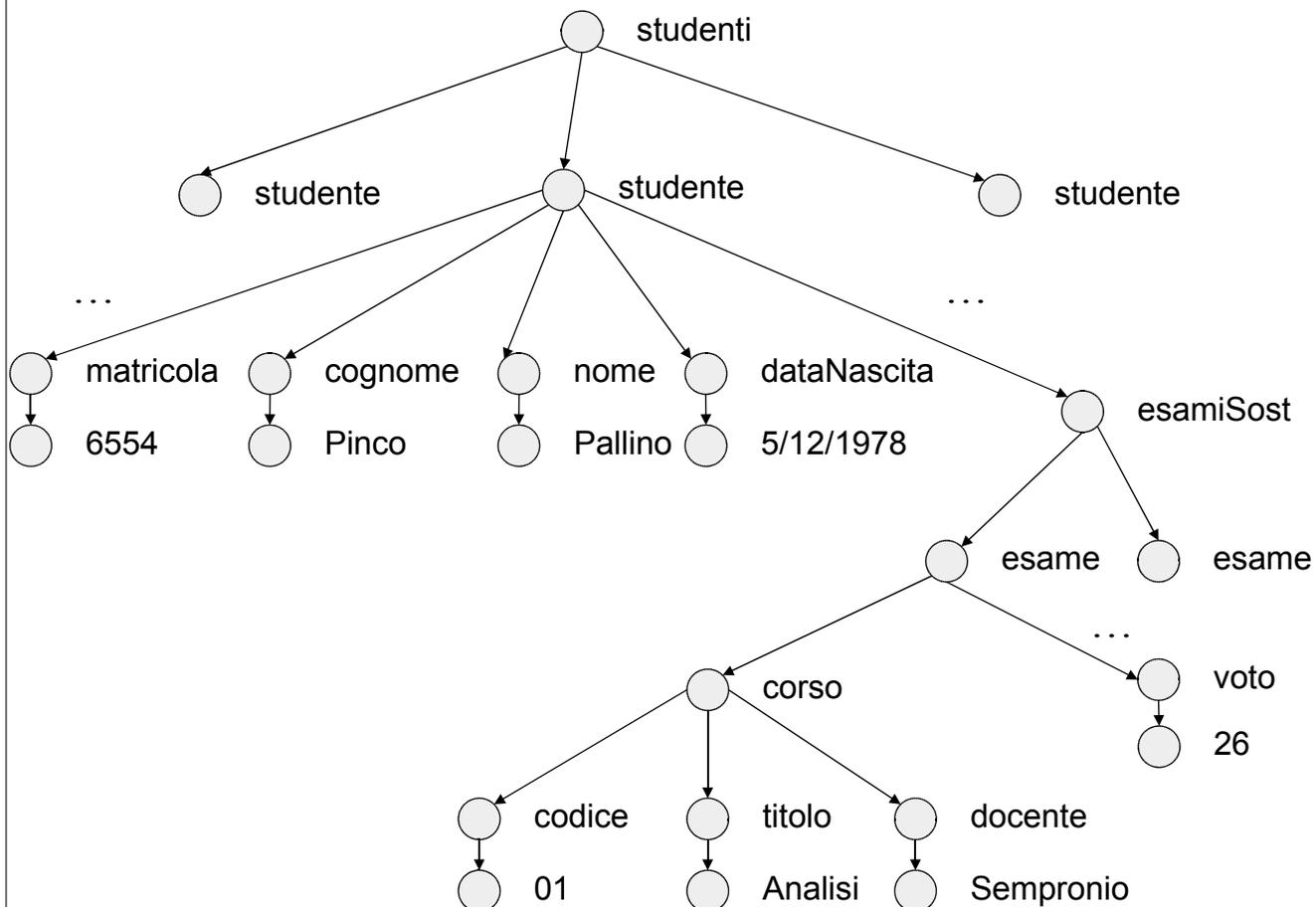
Modello concettuale dei dati

- ◆ 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



Rappresentazione 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>
</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 informazioni
- ◆ Ogni modello ha i suoi linguaggi
- ◆ DBMS relazionali: SQL, QBE
- ◆ DBMS a oggetti: OQL
- ◆ XML: XPath, XQuery

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