

# I dati

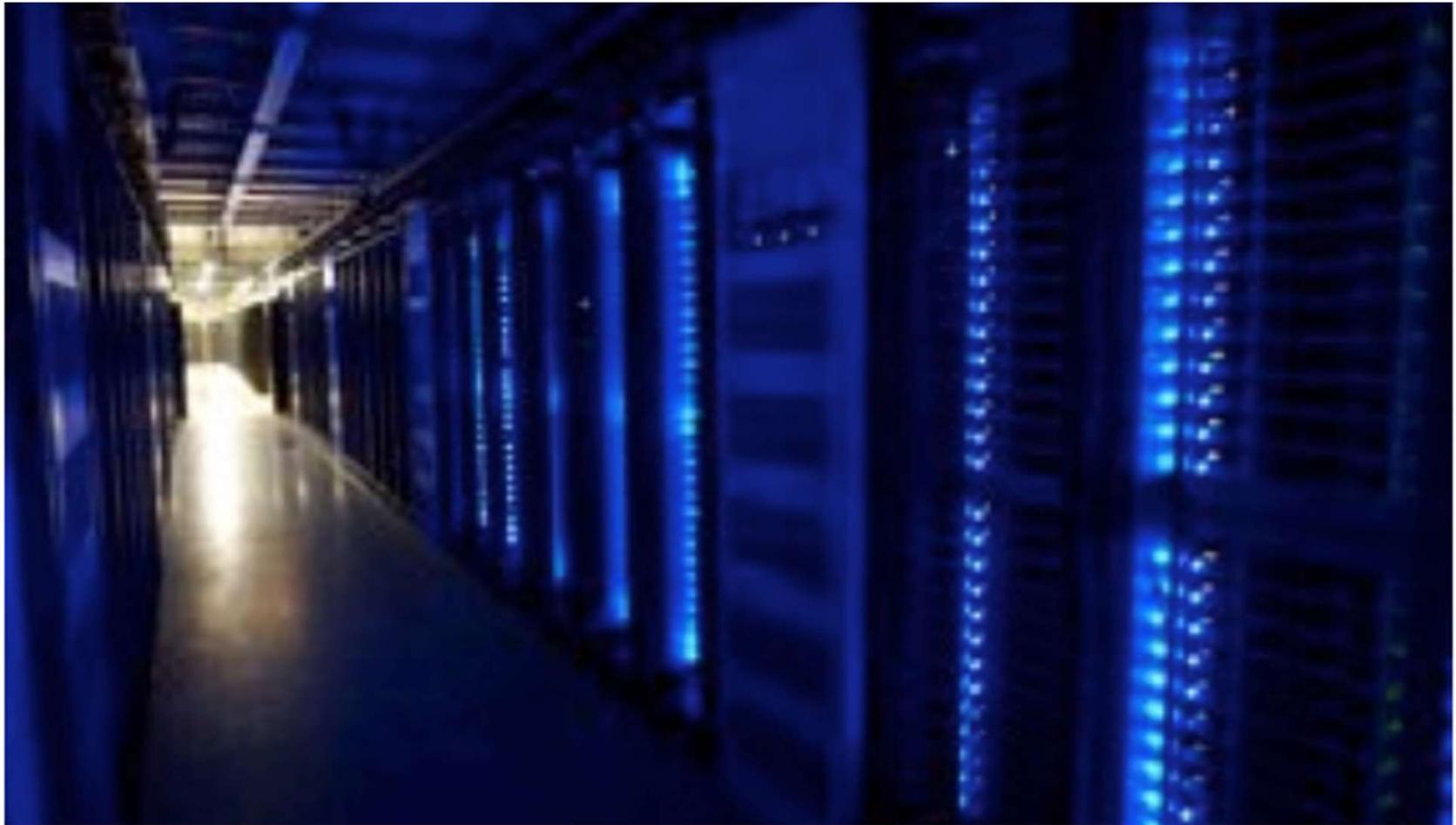
Giovanni Sartor

Giuseppe Contissa

# I dati nella memoria del computer

```
1010101011010110101001010101101010110101010101011010
1010101010110101010101101110101010101011010101010101
1010101101010101010101101010101010101101010101010110
1101010101010000101010110101001010101010101101010101
1010010100001010101010111001011001010010101011010100
0001101010100101001101000101010101101011010101010101
0011010101101010100100101010101010101010101011010101
1010101011010110101001010101101010110101010101011010
1010101010110101010101101110101010101011010101010101
10101011010101010101011010101010101101010101010110
1101010101010000101010110101001010101010101101010101
1010010100001010101010111001011001010010101011010100
0001101010100101001101000101010101101011010101010101
0011010101101010100100101010101010101010101011010101
1010101010110101010101101110101010101011010101010101
10101011010101010101011010101010101101010101010110
1101010101010000101010110101001010101010101101010101
1010010100001010101010111001011001010010101011010100
0001101010100101001101000101010101101011010101010101
0011010101101010100100101010101010101010101011010101
```

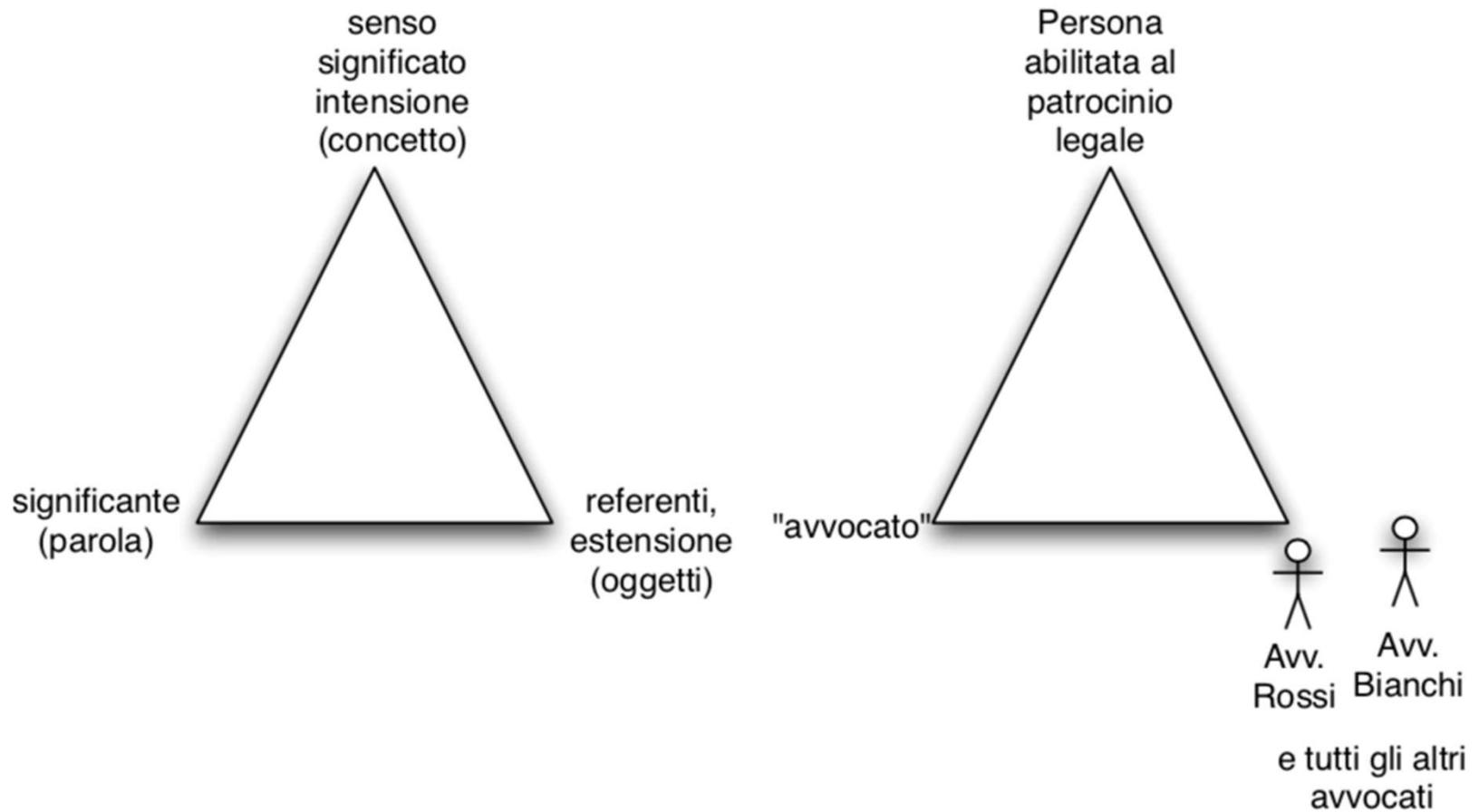
# Data centre di Facebook



# Dati e concettualizzazioni

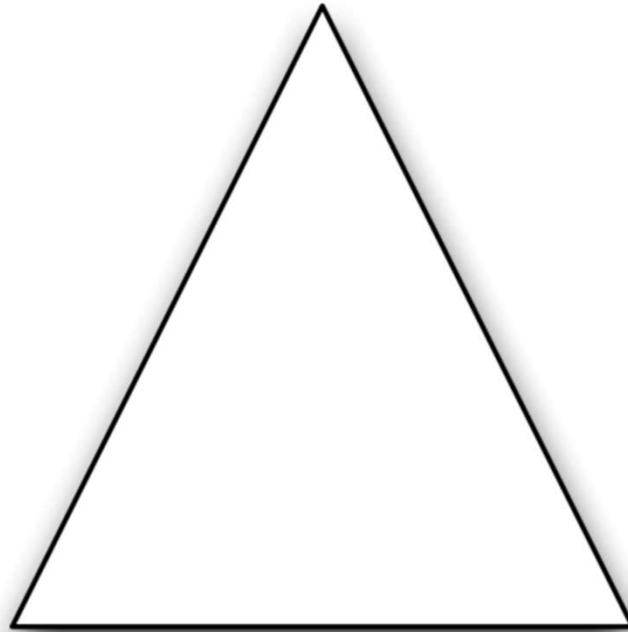
- la rappresentazione dei dati in un sistema informatico presuppone una concettualizzazione, cioè l'individuazione dei concetti da utilizzare per qualificare e ordinare la realtà rappresentata;
- ogni rappresentazione informatica presuppone una ontologia, cioè la determinazione di quali tipi di cose (entità) e di quali rapporti tra esse (associazioni) s'intendano rappresentare, e di quali informazioni su tali cose e rapporti si ritengano rilevanti.

# Triangolo semiotico



# Classe e istanze I

la classe "avvocato"



avvocato  
(il nome della  
classe)

Le istanze  
della classe  
(registrazioni  
concernenti singoli  
avvocati)

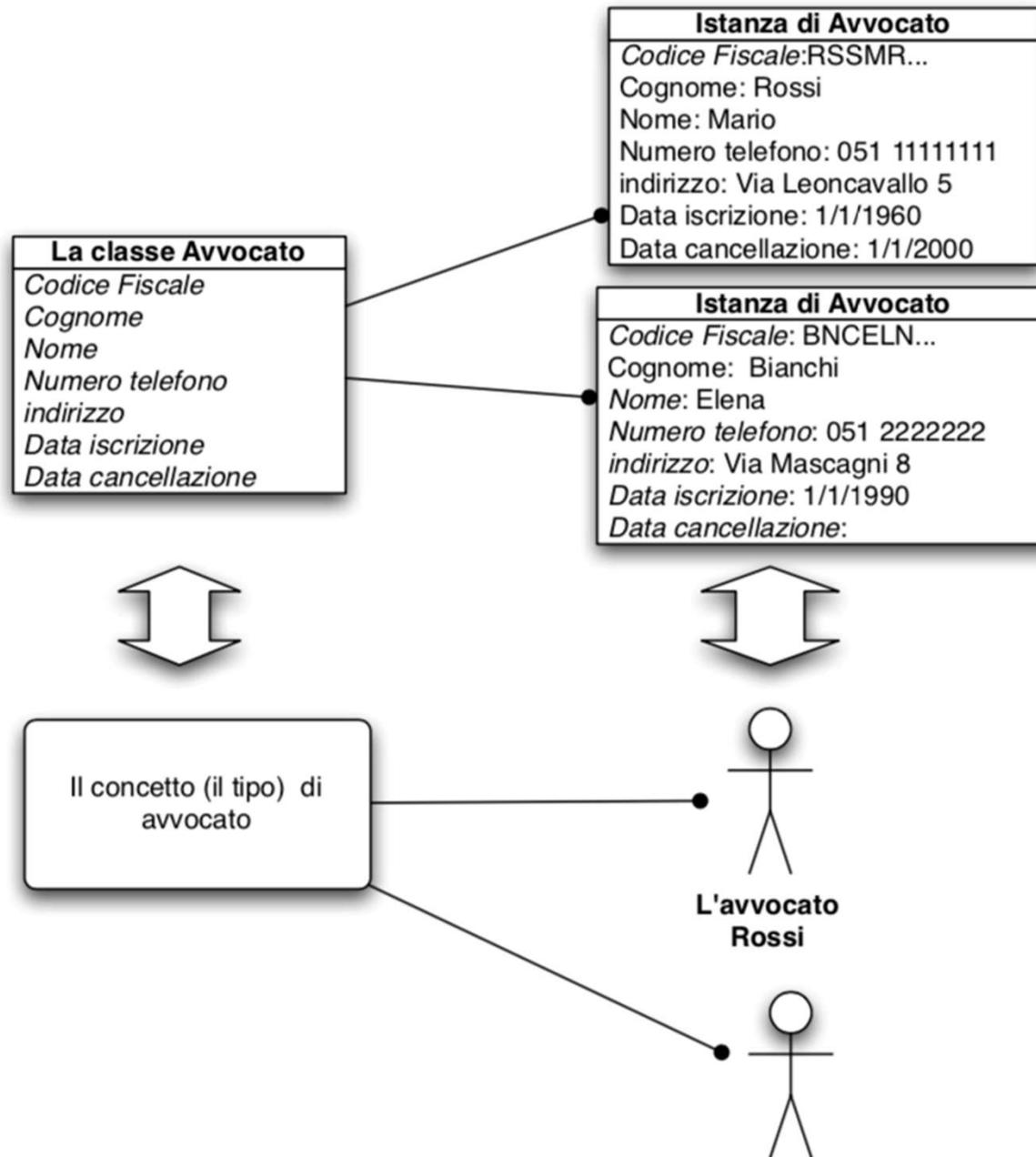
# Classi e attributi

- il termine “classe” è usato per indicare un concetto generale, vista come uno schema o struttura dalla quale si possono trarre molteplici istanze individuali, attinenti ad oggetti dello stesso tipo
- gli attributi di una classe sono le caratteristiche che possono essere attribuite alle istanze individuali della classe

# La scelta di classi e attributi

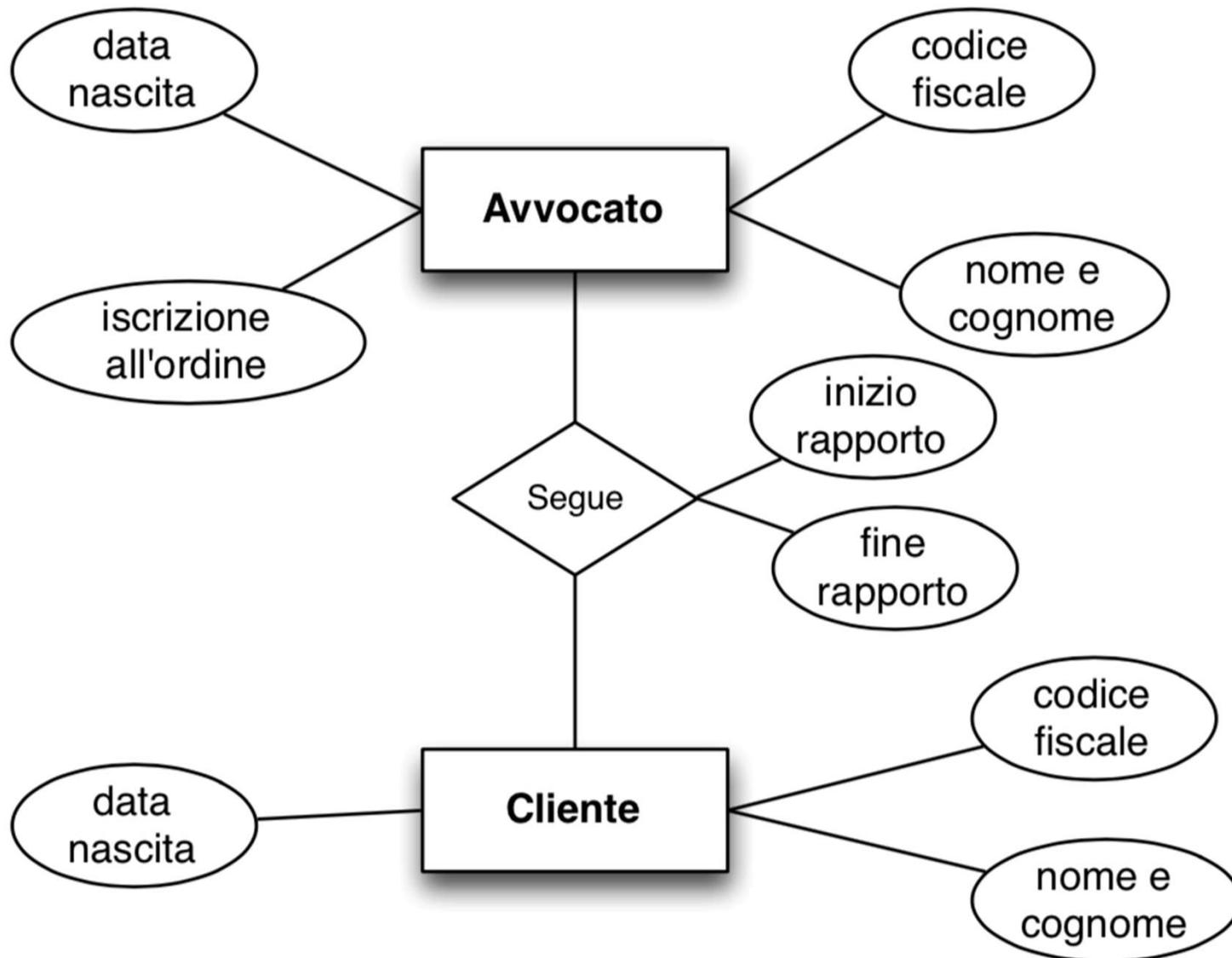
- Una buona rappresentazione informatica presuppone che si individuino le classi rilevanti per il dominio considerato e che esse siano caratterizzate da attributi che colgano gli aspetti significativi:
  - Chiave: attributo o una combinazione di attributi che individui in modo univoco uno e uno solo degli oggetti rientranti nella classe di riferimento (come il numero di codice fiscale per gli avvocati).
  - attributi sovrabbondanti rispetto allo scopo perseguito può condurre a violazioni della tutela dei dati, con particolare riferimento al requisito della necessità dei dati stessi

# Classe e istanze II



# Associazione

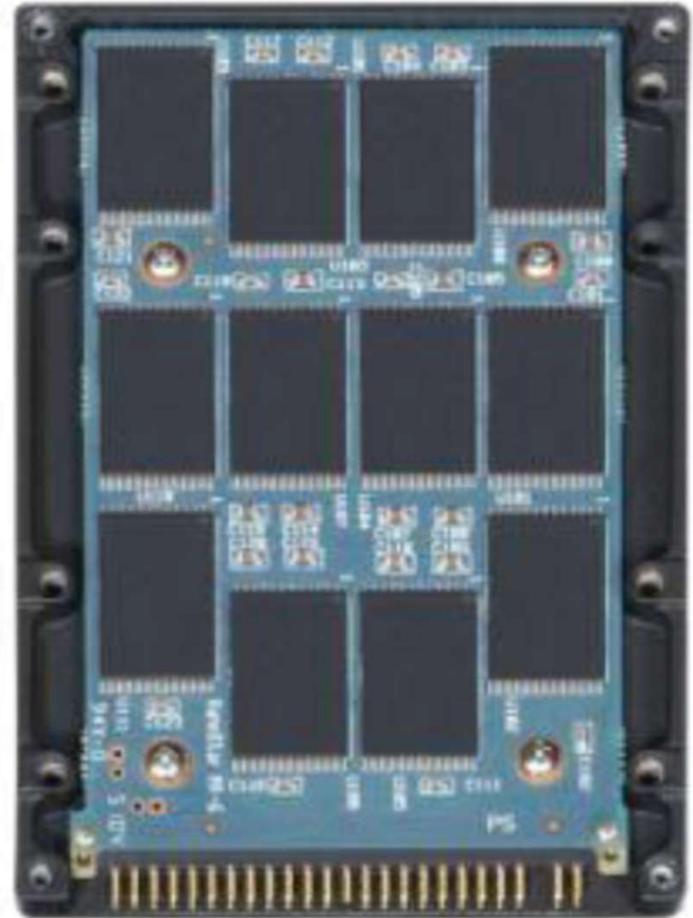
I rapporti tra entità sono rappresentati mediante associazioni.



# Dispositivi di memorizzazione



Traditional hard disk drive



Solid state hard drive

# L'organizzazione dei dati

- archivi: file strutturati, possibilmente raccolti in database
- testi: file non strutturati, possibilmente raccolti in sistemi di information retrieval

# Archivi

- Un archivio raccoglie le istanze di una certa classe (gli indirizzi, le anagrafiche studenti, i corsi di giurisprudenza, ecc.)
  - ad ogni istanza della classe è dedicato un record
  - ogni record è diviso in campi, ciascuno dei quali contiene il valore di un particolare attributo della classe
  - il formato di record indica la suddivisione del un record in determinati campi.

# Suddivisione in campi

- A lunghezza fissa

Mario Rossi +39051333333 Bologna Via Bianchi 7 ...

- Con separatori

Mario Rossi | | +39051333333 | | Bologna | | Via Bianchi 7 | | . . .

- Con etichette

<nome>Mario

Rossi</nome><telefono>+39051333333</telefono>

<citta>Bologna</citta><indirizzo>Via Bianchi 7</indirizzo>. . .

# File di testo

Consistono di componenti testuali (sequenze di parole), accompagnate da indicazioni sulla loro presentazione a video o a stampa, cioè da indicazioni per la formattazione, la visualizzazione, la strutturazione delle componenti testuali

- le indicazioni possono essere incomprensibili (come in Word e in generale nei sistemi WYSIWYG)
- possono essere comprensibili con difficoltà (con in RTF)
- possono essere esplicite, come in HTML o in LATEX

# Composizione finale

## **Informatica giuridica**

Giovanni Sartor  
1 Settembre, 2006

### **1 Introduzione**

L'informatica giuridica



# File in RTF

```
\b\f1\fs32\lang1033\langfe1033\kerning28\cgrid\langnp1033\langfenp1033 {\insrsid6384723
\charrsid6384723 Informatica giuridica
\par }\pard\plain \qc
\li0\ri0\nowidctlpar\tx560\tx1120\tx1680\tx2240\tx2800\tx3360\tx3920\tx4480
\tx5040\tx5600\tx6160\tx6720\fauto\rin0\lin0\itap0\pararsid6384723
\lang1033\langfe1033\cgrid \langnp1033\langfenp1033 {\insrsid6384723\charrsid6384723
Giovanni Sartor
\par 1 Settembre, 2006
\par }\pard\plain \s1\ql
\li0\ri0\sb240\sa60\keepn\widctlpar\aspalpha\aspnum\fauto\outlinelevel0
\adjustright\rin0\lin0\itap0\pararsid6384723 \b\f1\fs32\lang1033\langfe1033\kerning32\cgrid
\langnp1033\langfenp1033 {\insrsid6384723\charrsid6384723 1 Introduzione
\par }\pard\plain \ql
\li0\ri0\nowidctlpar\tx560\tx1120\tx1680\tx2240\tx2800\tx3360\tx3920\tx4480
\tx5040\tx5600\tx6160\tx6720\fauto\rin0\lin0\itap0\pararsid6384723
\lang1033\langfe1033\cgrid \langnp1033\langfenp1033 {\insrsid6384723\charrsid6384723
L\quote
informatica giuridica
\par }}
```

# File in Latex

```
\documentclass[12pt, a4paper]{article}  
\title{Informatica giuridica}  
\author{Giovanni Sartor}  
\begin{document}  
\maketitle  
\section{Introduzione}  
L'informatica giuridica  
\end{document}
```

# File in HTML

```
<html>  
<center><h1>Informatica giuridica</h1></center>  
<center>Giovanni Sartor</center>  
<center>1 Settembre 2006</center>  
<h2>Introduzione</h2>  
<p>L'informatica giuridica</p>  
</html>
```

# I formati di dati

Regole che stabiliscono come avviene la registrazione dei dati

- convenzioni per la rappresentazione binaria dei caratteri
- convenzioni per stabilire la funzione (la semantica) di ogni dato, nei file strutturati
- convenzioni per specificare la forma tipografica (nei file di testo)

I formati di dati sono un vincolo:

- Chi abbia accumulato una notevole quantità dei dati in un certo formato è vincolato all'utilizzo dell'unico software predisposto per quel formato, per non dover affrontare i costi e i disagi relativi al trasferimento in altro formato.
- Chi intenda condividere dati con altri soggetti è vincolato a predisporre quei dati in un formato accettato dai software impiegati da quei soggetti.

# Formati proprietari e formati aperti

- formati proprietari: sul loro uso e la loro evoluzione decide un soggetto privato, cioè, il soggetto che detiene il diritto d'autore sui software che producono i testi in quel formato (o sul formato o sugli algoritmi che lo producono)
- formati aperti: la loro definizione ed evoluzione è affidata a organismi imparziali, le loro specifiche sono rese pubblicamente disponibili, e possono quindi essere utilizzati da chiunque vi abbia interesse

# Basi di dati e sistemi di information retrieval

- basi di dati: consentono di aggregare, disaggregare, combinare, elaborare informazioni registrate su diversi archivi;
- Sistemi di IR: consentono di ricercare elementi testuali compresi all'interno di raccolte documentali.

# Basi di dati e DBMS

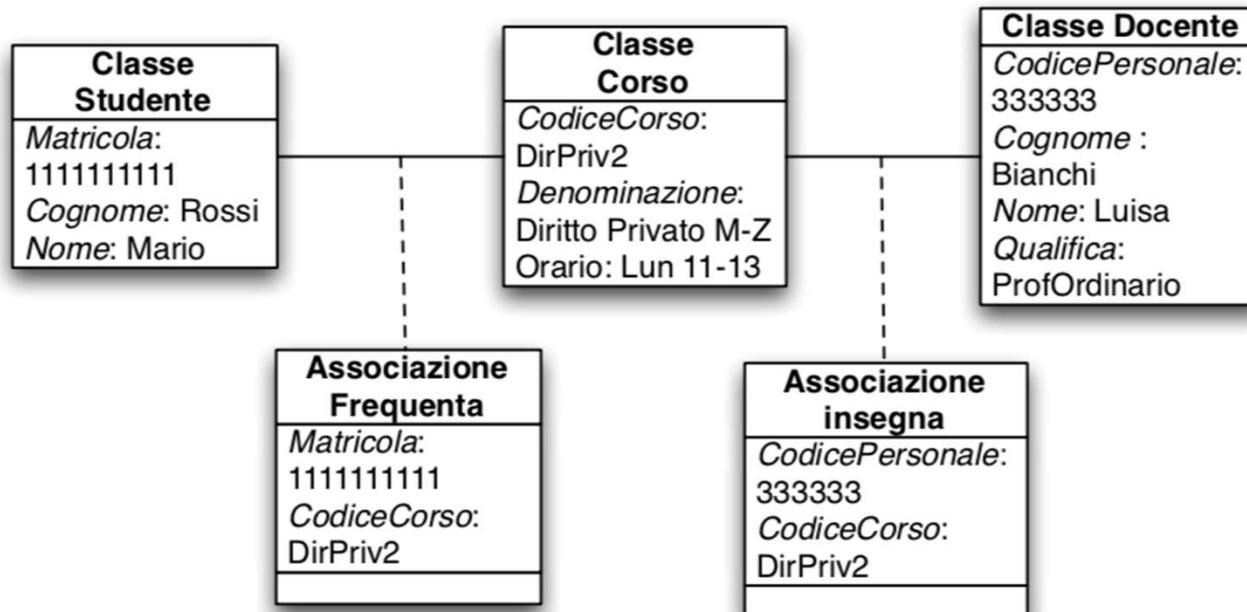
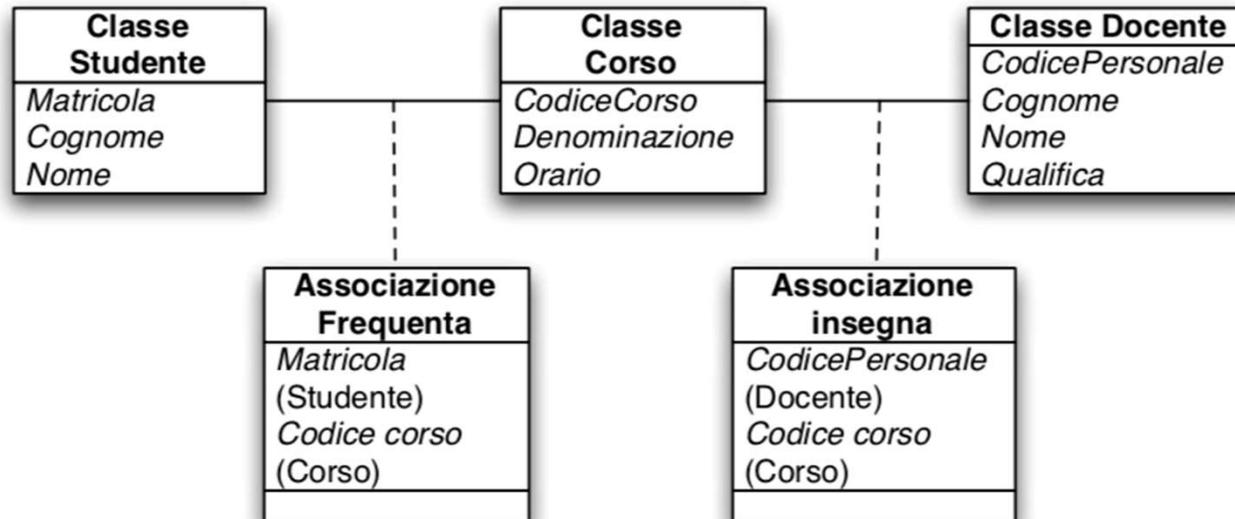
- Base di dati: banca di dati evolute, costituite da più archivi tra loro integrati, i cui dati sono elaborati in modo da offrire a diversi utenti le informazioni di cui hanno bisogno (e cui hanno diritto di accedere) nella forma che le renda meglio utilizzabili
- sistema per la gestione di basi di dati (*DataBase Management System* - DBMS): sistema informatiche che funge da intermediario tra gli archivi di dati e gli utilizzatori (umani o automatici) dei dati stessi. Consente di rappresentare i dati in modo indipendente dai programmi destinati a usare quei dati. A tal fine esso mantiene dei dizionari di dati (data dictionary). Si tratta di raccolte di metadati (dati intesi a descrivere altri dati) che specificano il significato e la struttura dei tipi di dati (classi) registrati nella base di dati, e le convenzioni usate nella loro registrazione

# DB e qualità dei dati

Un DBMS può contribuire ad assicurare la qualità dei dati stessi:

- *Coerenza dei dati.* Il DBMS permette a diverse applicazioni di utilizzare gli stessi dati, integrandoli e riducendo le duplicazioni (la ridondanza).
- *Integrità.* Il DBMS contribuisce a far sì che la base di dati sia integra, cioè contenga solo dati corretti.
- *Ripristino.* Il DBMS comprende procedure automatiche per il ripristino della base dei dati in caso di errori e malfunzionamenti imputabili all'hardware o al software.
- *Sicurezza.* Il DBMS controlla l'intera base di dati, e quindi può garantire che a essa si acceda solo secondo certe modalità e procedure, e dopo aver ottenuto idonee autorizzazioni (che possono essere diverse per diversi tipi di dati).
- *Riservatezza.* L'uso di un DBMS può facilitare il rispetto della riservatezza dei soggetti cui i dati afferiscono, specialmente quando si debbano trattare grandi quantità di dati, rendendoli accessibili a diversi utenti

# Modello di base di dati: classi e associazioni



# Dal modello alla base di dati relazionale

- ogni classe e ogni associazione è rappresentata mediante una distinta tabella (relazione);
- ogni riga della tabella rappresenta un'istanza della classe o associazione (ad esempio un determinato studente, o la frequentazione di un determinato corso da parte di un determinato studente);
- ogni colonna della tabella riporta i valori di un determinato attributo (ad esempio, la seconda colonna della tabella 11111111 riporta i valori dell'attributo Cognome);
- le associazioni tra diverse tabelle sono rappresentate affiancando le chiavi delle entità associate.

# Base di dati relazionale

**Relazione Studenti**

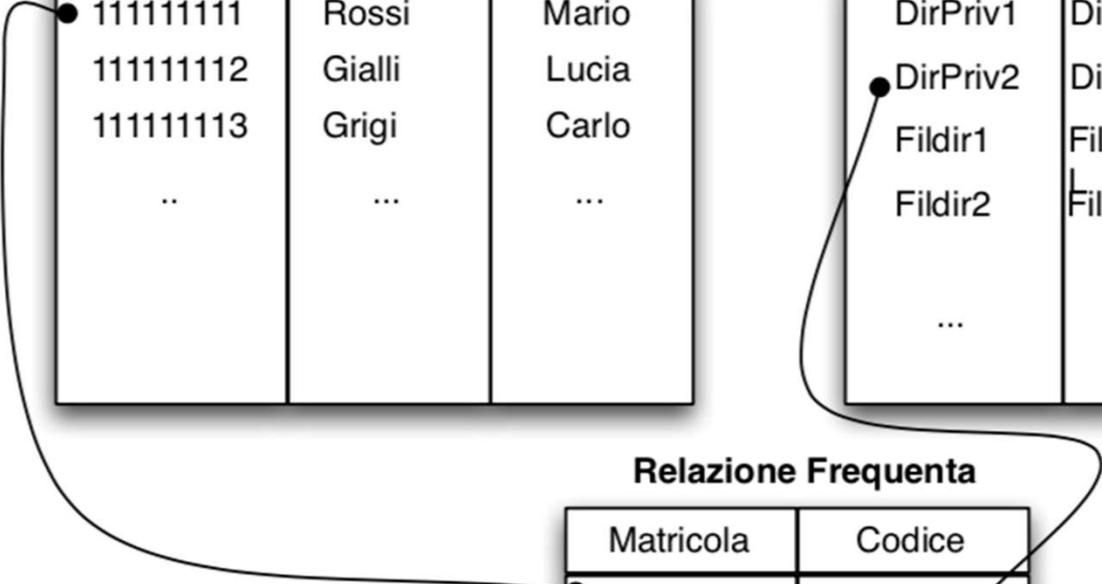
Matricola	Cognome	Nome
111111111	Rossi	Mario
111111112	Gialli	Lucia
111111113	Grigi	Carlo
..	...	...

**Relazione Corsi**

Codice	Denominazione	Orario
DirPriv1	Diritto privato A-L	Lun 11-13
DirPriv2	Diritto privato M-Z	Lun 11-13
Fildir1	Filosofia del diritto A-	Mar 10-11
Fildir2	Filosofia del diritto M-Z	Merc 9-10
...	...	...

**Relazione Frequenta**

Matricola	Codice
111111111	DirPriv2
111111111	FilDir2
111111112	DirPriv1
...	...



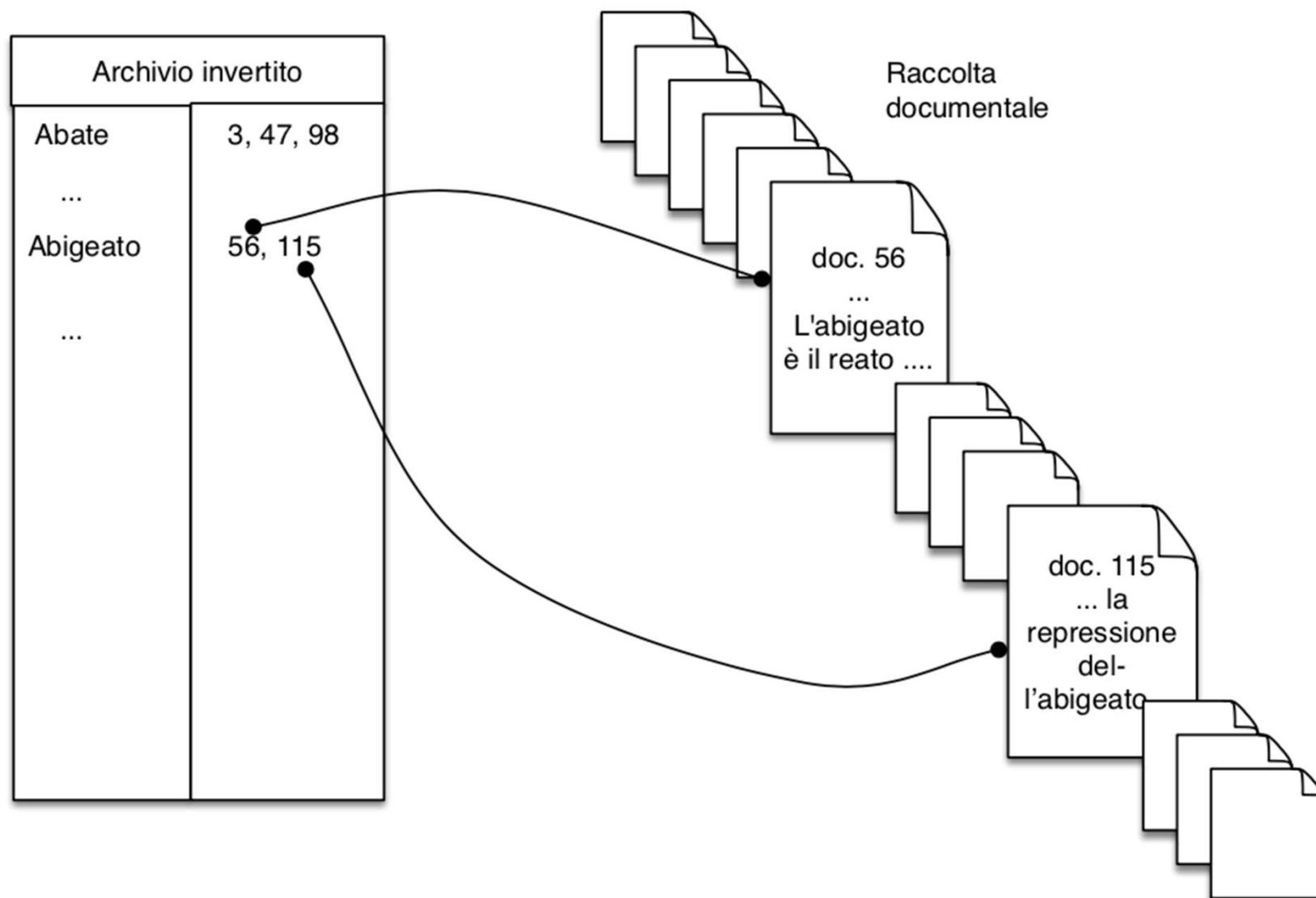
# Information retrieval

software per *l'information retrieval* (ricerca di informazioni): recupera (*retrieve*) le informazioni testuali estraendole da grandi masse documentali.

Sistema per la ricerca mediante indici:

- l'indicizzatore, per la creazione degli indici;
- il motore di ricerca (search engine) che, sulla base degli indici, individua ed estrae le informazioni richieste.

# File invertito



# Prestazioni dei sistemi per l'information retrieval

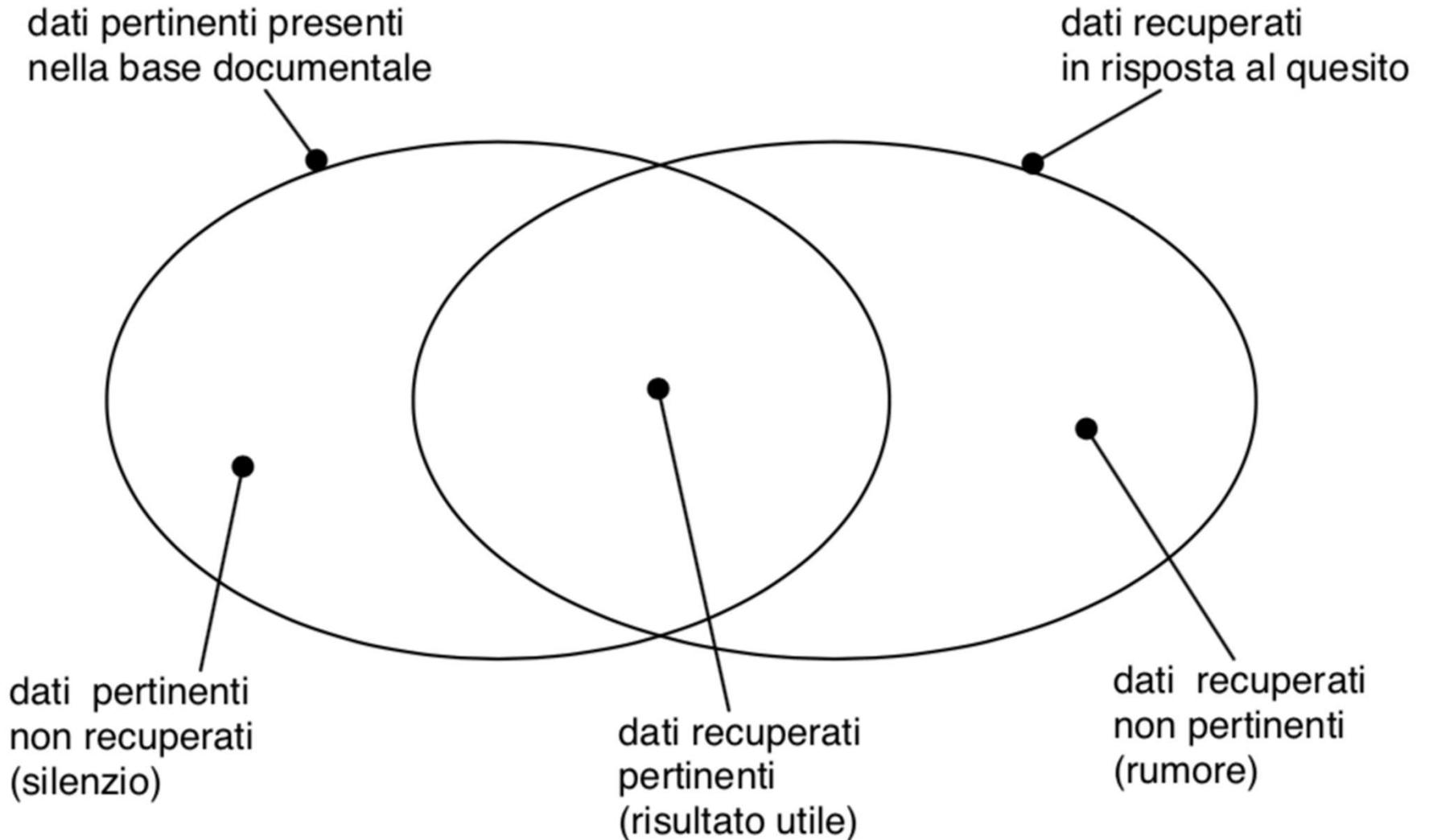
Richiamo e rumore:

- il *richiamo (recall)*, cioè la quantità di documenti pertinenti ritrovati, in rapporto a tutti i documenti pertinenti contenuti nella base documentale;
- la *precisione*, cioè la quantità di documenti pertinenti ritrovati, in rapporto a tutti i documenti recuperati.

Silenzio e rumore:

- il *silenzio* è il difetto nel richiamo, e risulta dalla relazione tra i documenti pertinenti non recuperati (i documenti sui quali il sistema è erroneamente silente), in rapporto a tutti i documenti pertinenti contenuti nel database.
- il *rumore* è il difetto nella precisione, e risulta dalla relazione tra la quantità di documenti recuperati non pertinenti (i documenti sui quali il sistema è erroneamente rumoroso), in rapporto a tutti i documenti recuperati.

# Silenzio e rumore



# L'operatore AND

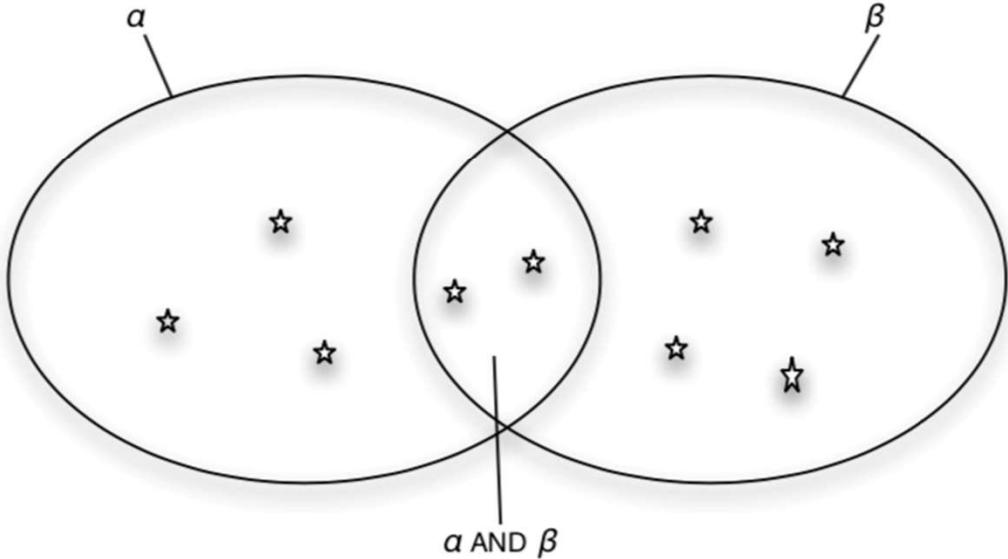
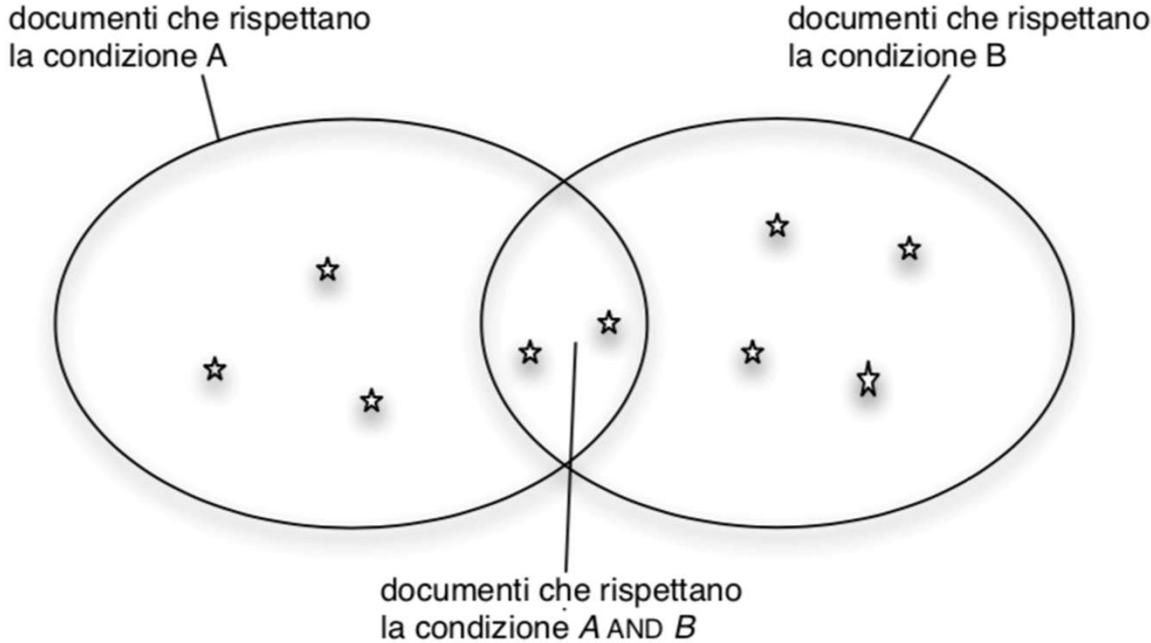
Domanda congiuntiva:

- Documento contiene “furto” AND Documento contiene “bestiame”

e in forma abbreviata

- furto AND bestiame

# Domanda congiuntiva



# Domanda congiuntiva complessa

- Titolo contiene “commercio” AND Articolato  
contiene “recesso” AND AnnoDiEmanazione = 1998

# L'operatore OR

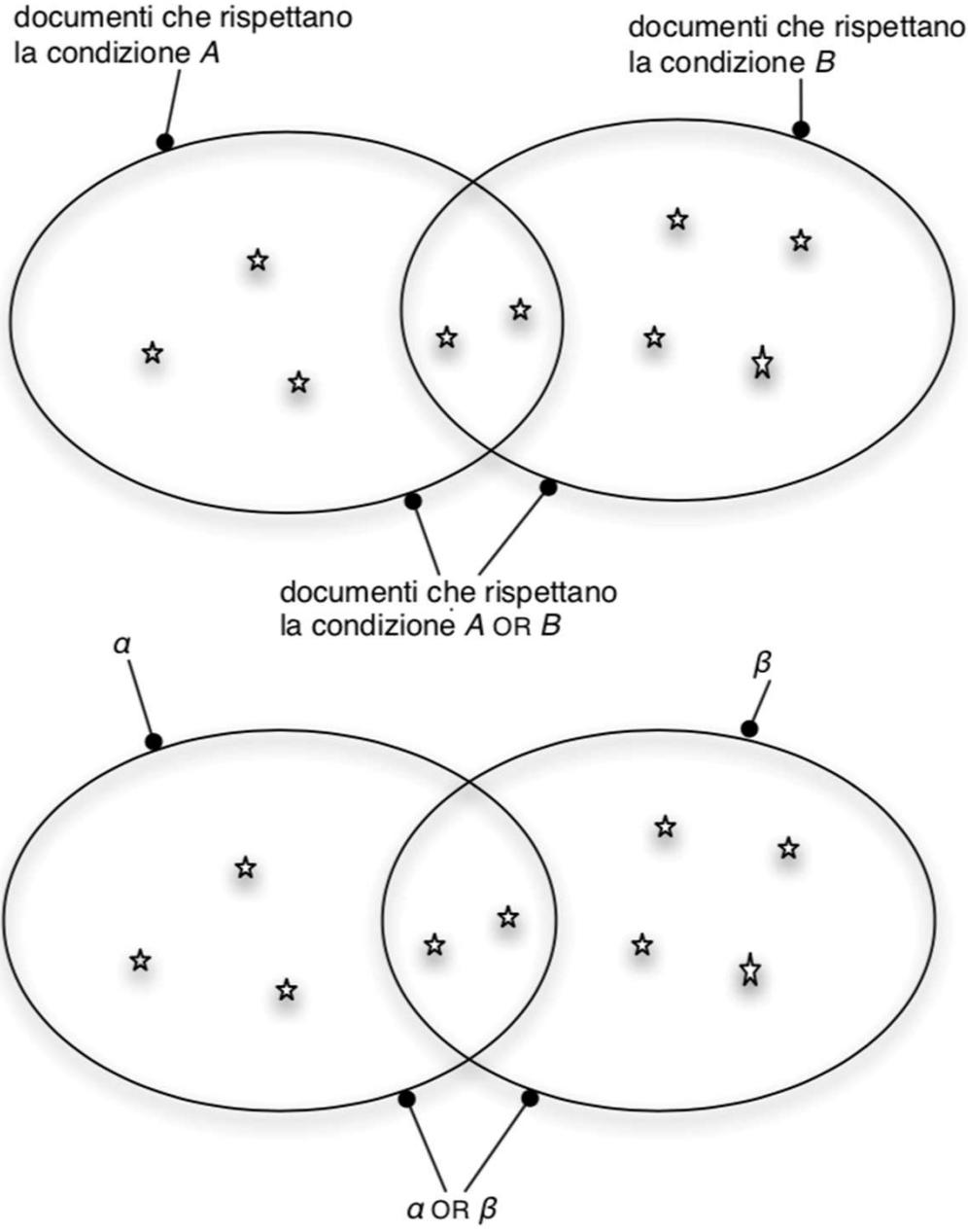
Domanda disgiuntiva:

- Documento contiene “furto” OR Documento contiene “bestiame”

e in forma abbreviata

- furto OR bestiame

# Domanda disgiuntiva



# L'operatore NOT

Domanda negativa:

- NOT Documento contiene “furto”  
e in forma abbreviata
- NOT furto

# Domanda negativa

