



## 2. L'ingegneria della conoscenza e il Web



- Riassunto della scorsa lezione
- Prospettiva storica
- L'ingegneria della conoscenza oggi
- Il Web Semantico
- I triangoli aristotelici
- Realtà e sua rappresentazione



II-1

2

## RIASSUNTO DELLA SCORSA LEZIONE

- Tipi di conoscenza
  - Dichiarative vs procedurali
    - Conoscere vs saper fare
  - Terminologiche
  - Nomologiche
  - Fattuali
  
- Fonti di conoscenza
  - Esperienza diretta
  - Ragionamento
    - Deduttivo, abduzione e induzione
  - Comunicazione
  - Memoria



II-1

3

- Il ruolo della conoscenza
  - Interpretare la realtà
  - Prevedere l'evoluzione della realtà
  - Agire in modo razionale modificando la realtà
  
- Storia dell'AI (in relazione al KE)
  - 55-65 Risoluzione automatica di problemi
  - 65-75 Rappresentazione simbolica della conoscenza
  - 75-85 Ingegneria della conoscenza e sistemi esperti
  - 85-95 Crisi e strade alternative: soft computing
  - 95- ... Web Semantico



II-1

4

## I sistemi esperti

- Con i metodi dell'ingegneria della conoscenza sono stati sviluppati diversi tipi di applicazioni d'interesse industriale.
- A partire dal 1975 e per tutti gli anni novacentottanta le applicazioni più diffuse sono stati i sistemi esperti.
  - Un sistema esperto è un sistema software in grado di risolvere problemi utilizzando conoscenze specialistiche;
  - applicazioni tipiche sono
    - l'analisi di grandi volumi di dati,
    - la diagnosi dei guasti di impianti industriali,
    - la diagnosi medica,
    - la configurazione di sistemi informatici complessi,
    - il progetto industriale,
    - la pianificazione della produzione
    - etc.



II-1

5

## Successi e insuccessi

- In certi casi l'adozione di sistemi esperti ha permesso di automatizzare attività complesse là dove tecniche informatiche più tradizionali avevano fallito.
- Tuttavia i sistemi esperti non sono mai diventati prodotti di grande diffusione, e questo per almeno due ragioni:
  - lo sviluppo di un sistema esperto è un'impresa lunga e costosa, che può essere affrontata soltanto per applicazioni di nicchia in cui sia possibile investire una notevole quantità di risorse;
  - i sistemi esperti richiedono una notevole competenza anche per l'uso e soprattutto per il mantenimento, e proprio perché si basano su conoscenze specialistiche non possono essere facilmente trasferiti da un'area applicativa a un'altra.



II-1

6

Oggi

- Oggi la tecnologia dei sistemi esperti continua a rivestire interesse in alcuni settori applicativi, ma è improbabile che possa dar luogo a prodotti di vasta diffusione.
- A partire dagli anni novanta ha cominciato a profilarsi un altro settore di grande interesse per l'ingegneria della conoscenza:
  - il web semantico (Antoniou, van Harmelen, 2004).
- Prima di occuparci di questo concetto tratteremo una breve descrizione di alcuni fattori che stanno influenzando il progresso dell'informatica.



II-1

7

Sviluppi recenti dell'informatica

- Da più di un decennio assistiamo a una significativa rivoluzione nel campo dell'informatica, cui le tecnologie del web stanno dando un notevole contributo.
- A grandi linee i punti centrali di questa rivoluzione sono:
  - la centralità dei dati,
  - l'interoperabilità delle applicazioni e
  - l'utilizzo delle tecnologie del web in reti di diverse tipologie.



11-1

8

## La centralità dei dati

- Qualsiasi applicazione informatica è costituita da
  - dati e
  - programmi;le due componenti non hanno però lo stesso peso in applicazioni di tipo differente.
- Consideriamo come primo esempio un'applicazione per la simulazione di sistemi dinamici.
  - In un caso del genere i dati si possono ridurre a un vettore che rappresenti lo stato iniziale del sistema e all'assegnamento di un valore a un piccolo numero di parametri del sistema.
  - Sulla base di questi dati è possibile condurre complesse simulazioni per indagare, ad esempio, l'evoluzione nel tempo dello stato del sistema.
  - Per ottenere risultati di qualità, naturalmente, sono estremamente critici gli algoritmi utilizzati per la simulazione;
  - ci troviamo quindi di fronte a un tipo di applicazione che potremmo denominare *centrata sul processo*.



11-1

9

## La centralità dei dati (2)

- Un esempio molto diverso è costituito dalle applicazioni che si fondano sull'utilizzo di basi di dati.
  - In questo caso l'informazione utile risiede essenzialmente nei dati, e i processi si limitano spesso a reperire i dati necessari e a fornirne una presentazione adeguata.
  - Le applicazioni di questo tipo possono essere denominate *centrate sui dati*.



II-1

10

### Centralità dei dati (3)

- Naturalmente non avrebbe senso chiedersi quale tipo di applicazione informatica sia più importante: a seconda delle necessità si utilizzerà un sistema centrato sul processo o centrato sui dati.
- È vero però che le applicazioni centrate sui dati sono sempre più diffuse in termini percentuali: lo stesso web si presenta essenzialmente come un colossale archivio di dati interrogato quotidianamente da milioni di utenti.
- Non è quindi scorretto affermare che negli ultimi anni l'interesse dell'informatica si è gradualmente spostato dai sistemi centrati sul processo ai sistemi centrati sui dati.



II-1

11

### L'interoperabilità "interna"

- Per molti anni le aziende hanno sviluppato e utilizzato applicazioni "chiuse", che operano su dati per lo più in formato proprietario.
- Questo tipo di applicazioni crea problemi già all'interno della singola azienda, perché in una situazione del genere è difficile far interagire applicazioni distinte; a maggior ragione è arduo combinare applicazioni appartenenti ad aziende diverse.
- All'interno di una singola azienda è possibile raggiungere un buon livello d'interoperabilità utilizzando le tecnologie basate su XML; in particolare le tecnologie del web basate su XML consentono di sviluppare intranet aziendali integrando dati e servizi web (*web services*, WS) rivolti verso l'interno dell'azienda.
- Quando però si ha l'esigenza di superare i confini di un'intranet queste tecnologie si rivelano spesso insufficienti.



II-1

12

## L'interoperabilità "esterna" e i sistemi aperti

- Da qualche anno è nato un forte interesse per le applicazioni "aperte",
- in grado cioè di interoperare con applicazioni appartenenti ad aziende diverse
  - all'interno di un'extranet che connetta aziende federate o
  - con applicazioni comunque distribuite su internet.
- I servizi web possono essere utilizzati anche all'interno di un'intranet o di un'extranet, ma è chiaro che le loro maggiori potenzialità si rivelano al livello di internet.

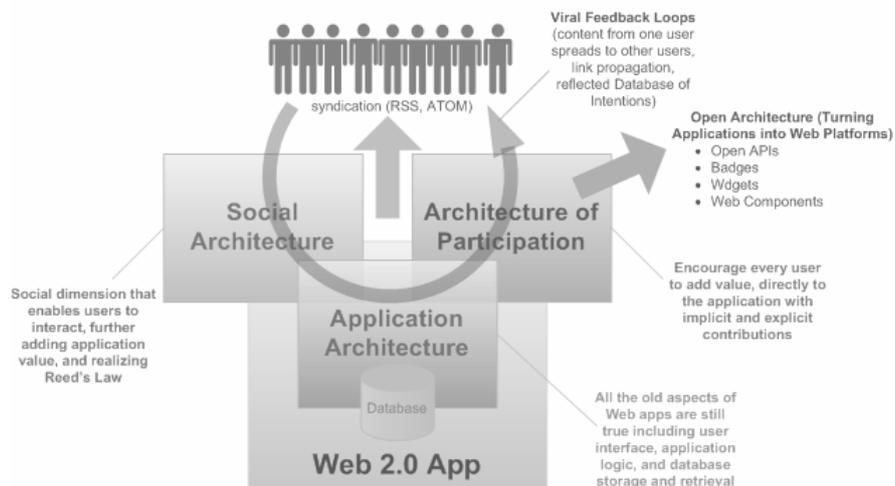


II-1

13

## Il Web 2.0 come esempio

**Web 2.0 Apps: Networked Applications that Explicitly Leverage Network Effects via an Architecture of Participation (optionally) reinforced by a Social Architecture**





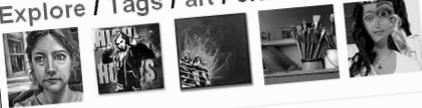
11-1  
14

## Social Tagging & Folksonomies

### Explore Flickr Through Tags

**art** australia baby beach birthday blue bw california cameraphone canada  
 canon cat chicago china christmas city dog england europe family flower  
 flowers food france friends fun germany holiday india italy japan london  
 me mexico music nature new newyork night nikon nyc paris park party  
 people portrait sanfrancisco sky snow spain summer sunset taiwan tokyo travel  
 trip usa vacation water wedding winter

### Explore / Tags / art / clusters



painting, graffiti, red, street, color, streetart,  
 blue, wall, urban, portrait

→ See more in this cluster...



sculpture, museum, architecture, statue,  
 gallery, building, modern, metal, sky, bronze

→ See more in this cluster...



nyc, newyork, newyorkcity, moma,  
 manhattan

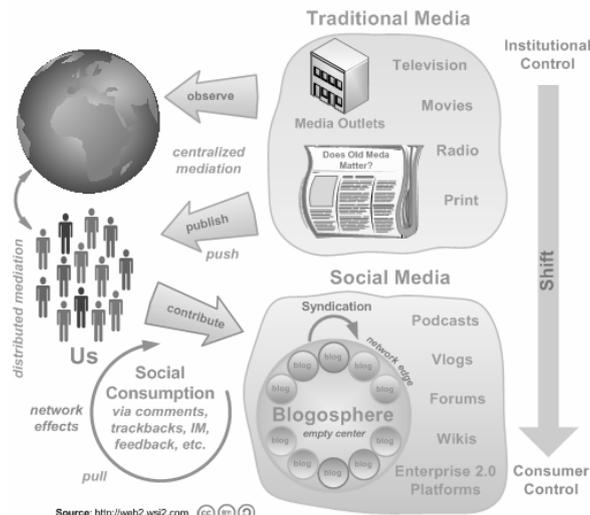
→ See more in this cluster...



11-1  
15

## Social Media

### The Emergence and Rise of Mass Social Media





II-1

16

## Il web semantico

- Come è noto, l'invenzione del web è attribuita a Tim Berners-Lee. Nella concezione di Berners-Lee, tuttavia, il web come lo conosciamo oggi è solo il primo gradino di una scala che prevede al gradino successivo il *web semantico*.
- L'idea è semplice:
  - *perché possano essere elaborati in modo completamente automatico, non è sufficiente che i dati sul web abbiano una struttura chiaramente definita (come è possibile fare utilizzando XML): è anche necessario che i dati abbiano un significato condiviso dalla comunità che li utilizza.*
- Per consentire lo sviluppo del web in questa direzione il W3C ha proposto e sta tuttora elaborando un insieme di tecnologie standardizzate: RDF, RDFS, OWL, ...



II-1

17

## Il Triangolo Aristotelico

- L'idea del web semantico è dunque centrata sull'assegnamento di un *significato* ai dati presenti sul web.
- Il concetto di significato, a sua volta, chiama in causa il rapporto fra *linguaggio* e *realtà*.
- Come tutti sappiamo, molti termini del linguaggio corrispondono a entità del mondo reale:
  - ad esempio, il termine "seggiola" (un sostantivo dell'italiano) può corrispondere
    - all'insieme di tutte le seggiole, oppure
    - a una seggiola specifica (un individuo membro dell'insieme delle seggiole).
- La relazione fra linguaggio e realtà (e quindi fra il termine "seggiola" e l'insieme delle seggiole concrete) è detta *relazione semantica*.



## Il Triangolo Aristotelico (2)

- Secondo una tradizione antica (che nella storia del pensiero occidentale si può far risalire ad Aristotele), il rapporto fra linguaggio e realtà non è diretto, ma è mediato dai *concetti*:
  - chi parla l'italiano riesce a collegare la parola "segiola" alle seggiole reali proprio perché possiede un concetto di seggiola;
  - In altre parole, la relazione fra linguaggio e realtà è mediata dalla *mente*.
- Più precisamente, dobbiamo distinguere fra:
  - il *termine* "segiola" dell'italiano ("chair" in inglese e così via);
  - l'*insieme di tutte le seggiole* che esistono nella realtà (l'*estensione* del termine "segiola");
  - il *concetto* di seggiola così come è rappresentato nella mente degli esseri umani (l'*intensione* del termine "segiola").



## Il Triangolo Aristotelico (3)

- La relazione fra queste entità è schematizzata in triangoli





11-1  
20

## Abusi linguistici

- Il gergo dell'ingegneria della conoscenza spesso confonde i tre vertici del triangolo:
  - una definizione di "segiola" sarà indifferentemente denominata "termine", "concetto" o "classe".
- Questo abuso linguistico è innocuo purché si tenga in mente la distinzione appena introdotta;
- a rigore, comunque, nei sistemi informatici sarebbe preferibile parlare sempre di *termini*, dato che un gli oggetti informatici sono comunque simboli appartenenti a un linguaggio formale.



11-1  
21

## Realtà e sua rappresentazione

- Il rapporto tra linguaggio ed immagine, ovvero tra rappresentazioni logiche ed analogiche, è un tema sul quale Magritte gioca con grande intelligenza ed ironia.
- Esiste un confine, non sempre coscientemente chiaro, tra realtà e sua rappresentazione.





In questo caso, guardando l'immagine di una pipa e leggendo la scritta sottostante che dice: "questa non è una pipa", la prima reazione è di chiedersi: "ma allora, cosa è?".

Il sottile inganno si svela ben presto, se si riflette che si sta guardando solo un'immagine, non l'oggetto reale che noi chiamiamo "pipa".



René Magritte, Questa non è una pipa, 1948