

Ing. Federico Chesani

SEMANTIC WEB

Outline

- Introduzione a Semantic Web
 - Caratterizzazione del World Wide Web odierno
 - La proposta Semantic Web
 - Architettura del Semantic Web
- Informazione semantica (I): RDF e RDFS
- Informazione semantica (II): OWL
- Logiche, proof e trust
- SW oggi: problemi e critiche

Il Web oggi (Web 1.0) ...

- I contenuti sono pubblicati su web per essere "human-readable"
 - Standard HTML concentrato su "come" rappresentare i contenuti
 - Nessun meccanismo per definire "cosa" si sta rappresentando
 - Alcuni tags (e.g. <title>) danno semantica implicitamente ma ...
 - ... il loro contenuto non è strutturato
 - ... il loro uso non è standardizzato

Il Web oggi (Web 1.0) ...

Cambiava tutto fatto, invece approvato un solo articolo. All'origine del blocco il voto di Mastella all'art. 91

Finanziaria, il voto slitta a domani Al Senato va in scena l'assurdo

Dini ripete: "Mani libere fino all'ultimo". Predicioneista
di CLAUDIO FUSANI



Lamberto Dini
ROMA. - L'"assurdo" è andato in scena oggi nell'aula di palazzo Madama sfidando la logica, il buon senso, la curiosità gelosum della politica e lasciandosi il peso alle monovoci di palazzo. Doveva essere il giorno dell'antesima votazione finale alla Finanziaria. Alle nove e mezzo del mattino, quando il presidente dell'Senato Franco Marini riprende i lavori mancano sette articoli - dal 91 al 97 - e circa ottanta votazioni.

Poch'ore ad è finita. Alle otto di sera, invece, l'articolo 91 è stato accantonato - rimasto - il 92 approvato, il 93 in piena discussione. E anche molto vivace. Si è no sono state fatte una decina di

votazioni. Un impegno assurdo, kakistico, che racconta una delle pagine più incredibili di questa

Possiamo identificare il titolo in base alla sua presentazione (<h1>,) ...

... se domani il grafico cambia il formato esterno, il nostro programma non funziona più!!!

```
<h1>
<! -- inizio TITOLO -->
<B>Finanziaria, il voto slitta a domani <br>
    Al Senato va in scena l'assurdo</B>
<! -- fine TITOLO -->
</h1>
```

Il Web oggi (Web 1.0) ...

- Le pagine contengono *collegamenti* (links) ad altre pagine
 - Nessuna informazione sul link ...
 - ... cosa rappresenta un certo collegamento?
 - ... cosa rappresenta la pagina/risorsa riferita da un link?

Il Web oggi (Web 1.0) ...

Actual Web = Layout + Routing

Problema: non è possibile ragionare in modo *automatico* sui dati

II Web oggi (Web 1.0) ...

- Web inteso come colossale *archivio di dati*, interrogato ogni giorno da milioni di utenti
 - L'accesso a tale archivio avviene tramite *motori di ricerca* ed in base a *keywords* ...
 - ... il successo di tale ricerca dipende da molti fattori
 - Bontà dell'algoritmo di ricerca
 - Numero delle pagine indicizzate
 - Contenuti/meta-contenuti delle pagine indicizzate

II Web oggi (Web 1.0) ...

- Il web è *universale*
 - Qualunque pagina può contenere collegamenti ad altre pagine
 - Approssimativamente, chiunque può pubblicare su web riguardo qualunque argomento
 - *Decentralizzazione* delle informazioni
 - *Inconsistenza* delle informazioni
 - *Incompletezza* delle informazioni

Semantic Web

Obiettivo: poter "utilizzare"/"ragionare su" la
enorme quantità di dati disponibili sul web in
maniera **automatica**

Come? **Estendendo** l'attuale web con
informazioni sul contenuto (informazioni
semantiche)

Proposto da Tim Berners-Lee nel 2001, in un articolo
pubblicato su *Scientific American*

Semantic Web

Principi che si vogliono salvaguardare:

- Universalità del web
- De-centralizzazione dell'informazione
- Inconsistenza dell'informazione
 - Inconsistenza dei contenuti
 - Inconsistenza delle informazioni di routing (links)
- Incompletezza dell'informazione

Aggiungere informazioni sul contenuto

Aggiungere informazioni sul contenuto non è sufficiente

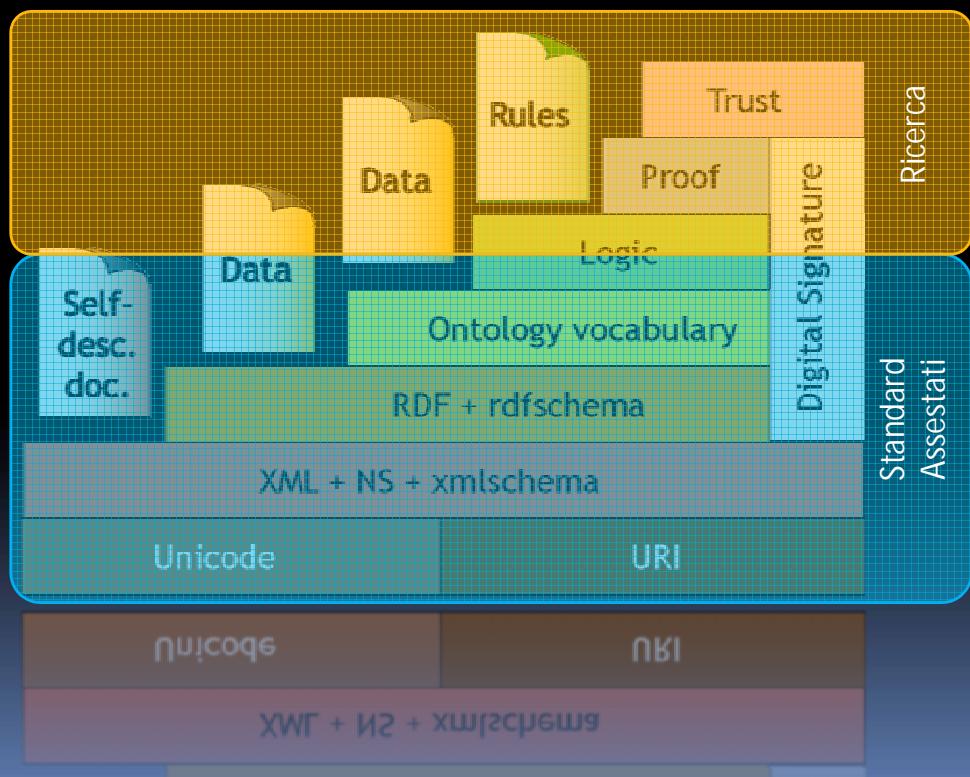
- Tali informazioni devono essere opportunamente *strutturate* (e.g., classificazione di Linneo per il mondo animale)
 - *Ontologie?*
- E' necessario aggiungere anche *regole di inferenza* (e.g., sillogismo aristotelico)
 - *Logica?*
- Applicando regole a dati si ottengono nuove conoscenze
 - Processo di prova (*proof*) con cui si dimostrano le conclusioni

Proof e Trust

Le "proofs" usate per inferire nuove informazioni possono essere scambiate ed analizzate per...

- ... giustificare i risultati ottenuti
- ... superare l'aspetto definitorio dell'IT
- ... ragionare ulteriormente sulla "fiducia" riguardo i risultati ottenuti
 - Le informazioni sul web sono (e saranno) incoerenti, contradditorie ed incomplete

Architettura del Semantic Web



SW – Applicazioni ?

Qualche esempio ...

- Migliorare i sistemi di ricerca (e.g., Portale Vodafone)
- Assistente remoto
- Integrazione Database (ambito industriale)
- Biblioteche ed archivi digitali (XMP Adobe)
- Semantic Web Services

Identificazione dei concetti

- In SW, i concetti vengono identificati tramite URI (*Uniform Resource Identifier*)
 - Gli URI sono per definizione un sistema di nomi unici
 - Ad ogni URI corrisponde uno ed un solo concetto...
 - ... però più URI possono definire lo stesso concetto!
 - URI (non URL): non è necessario che vi corrisponda un contenuto

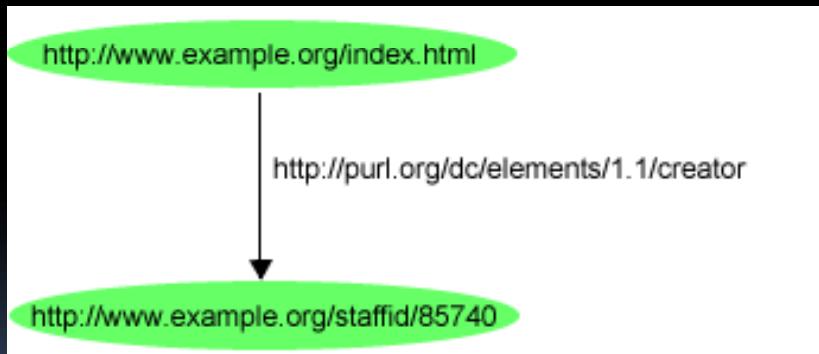
Esempi di URI:
http://www.repubblica.it
federico.chesani@unibo.it
ISBN 88-7750-483-8

Resource Description Framework (RDF/RDFS)

- Standard W3C
- Linguaggio (XML based) per la rappresentazione di conoscenza
- Obiettivo: fornire uno strumento *minimalista*
- Basato sul concetto di tripla:
 - < soggetto, predi cato, oggetto >
 - < ri sorsa, attri buto, val ore >
- Diverse rappresentazioni possibili (N3, Grafo, RDF/XML)

RDF – Rappresentazione a grafo

- Un nodo per il soggetto
- Un nodo per l'oggetto
- Un arco etichettato per il predicato



`http://www.example.org/index.html` has a **creator** whose value is `John Smith`

RDF – Rappresentazioni



RDF – Rappresentazioni

```
<rdf:RDF  
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
    xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#"  
>  
  
<contact:Person  
    rdf:about="http://www.w3.org/People/EM/contact#me">  
    <contact:fullName>Eric Miller</contact:fullName>  
    <contact:mailbox rdf:resource="mailto:em@w3.org"/>  
    <contact:personalTitle>Dr.</contact:personalTitle>  
</contact:Person>  
</rdf:RDF>
```

Posso chiedere al computer la mailbox di Eric Miller, senza sapere a priori se usa una casella di posta elettronica, piuttosto che una casella postale fisica, o quant'altro...
... e se Eric Miller cambia mailbox, il risultato della mia ricerca sarà sempre coerente!

RDF – Capacità espressive

Supporta:

- tipizzazione tramite attributo *type* (che assume come valore un URI)
- soggetto/oggetto di una frase possono essere anche *contenitori* (*bag*, *sequence*, *alternati* ve)
- *meta-asserzioni*, tramite *reificazione* delle asserzioni (“Marco afferma che Federico è autore di una certa pagina web”)

RDF Schema

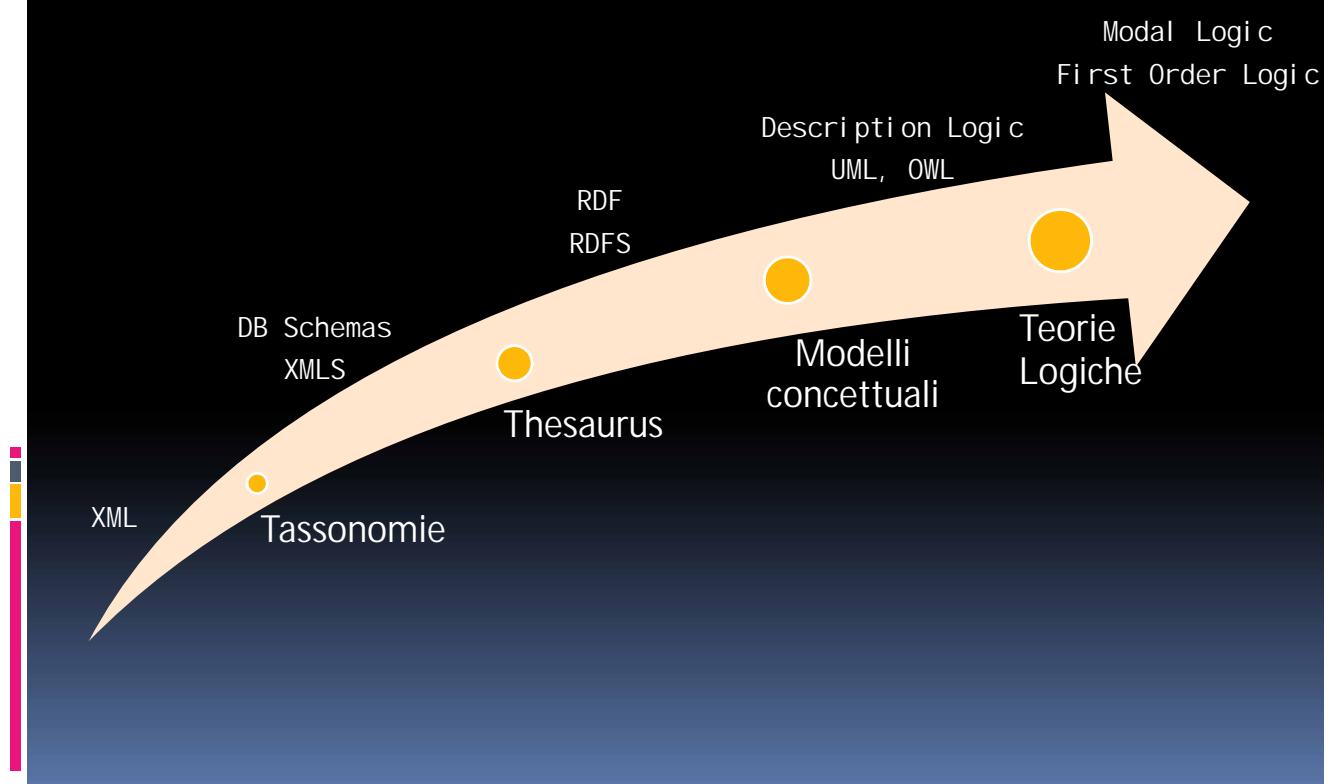
- RDF può essere inteso come descrizione di attributi di risorse e dei valori assunti da tali attributi
- RDFS permette di descrivere classi, proprietà e relazioni con altre risorse (simile ad OO)
 - *type*
 - *subClassOf*
 - *subPropertyOf*
 - *range*
 - *domain*

RDF e RDFS - Limiti

- Limitato potere espressivo
- Semantica non "ben definita"
- Non permette ancora di effettuare inferenze sulla conoscenza
- Non risolve l'ambiguità di più URI che si riferiscono ad uno stesso concetto

Soluzione proposta: **Ontology Web Language (OWL)**

Modelli semantici



Modelli semantici

- **Tassonomia**: insieme di termini organizzati in una gerarchia
 - rappresenta il fatto che esistono delle relazioni fra termini ...
 - ... ma non permette di stabilire la natura di tali relazioni
 - tipicamente gerarchie di tipo padre/figlio
 - la ricerca di un termine ha successo solo se si conosce a priori "dove" cercare

Modelli semantici

- **Thesaurus**: insieme di termini fra cui esistono più relazioni
 - oltre alla relazione padre/figlio (gerarchia), vi sono ulteriori relazioni associative (e.g. sinonimia, iperonimia, iponimia, olonimia, meronimia)
 - risolvono problemi tipici del linguaggio naturale, quali ambiguità e rindondanza
 - Esempio: WordNet

Modelli semantici

- **Modello concettuale**: è il modello di una particolare area di conoscenza (*dominio*). Definisce:
 - Entità del dominio
 - Relazioni fra le entità (in forma di proprietà ed attributi)
 - Regole su classi, attributi e relazioni
 - Regole di inferenza ... *Teorie Logiche!!!*

Ontologie – una definizione

Una ontologia è una **descrizione formale esplicita** di un **dominio** di interesse

- Classi
- Relazioni semantiche tra classi
- Proprietà associate ad un concetto (eventuali restrizioni)
- Eventuale livello logico (assiomi, regole di inferenza)

Ontologie – un esempio



Ontology Web Language (OWL)

- Standard W3C
- Si basa ed estende RDF/RDFS
- Dotato di semantica formale (supporta *Description Logic*)
- Tre livelli di espressività/complessità
 - OWL Lite (semplice, poco espressivo)
 - OWL DL (complessità della description logic, mediamente espressivo)
 - OWL Full (altamente espressivo ed oltre FOL, non decidibile)

OWL – caratteristiche

- **Classi:** subClassOf, intersectionOf, unionOf, complementOf, enumeration, equivalence, disjoint
- **Proprietà:** symmetric, transitive, functional, inverseFunctional, range, domain, subPropertyOf, inverseOf, equivalentProperty
- sulle **Istanze:** sameIndividualAs, differentFrom, allDifferent

Interesse per le ontologie

- Una ontologia fornisce un modello/strutturazione di un dominio
 - Risolve ambiguità di termini
 - Chiarifica/semplica le peculiarità di un dominio
 - Come ricaduta, analisi approfondita di un dominio/processo ...
 - ... elevato valore in ambito industriale/business

Ontologie – problemi aperti

- Equivalenza tra ontologie diverse
 - Ontology matching problem
 - Una ditta vende un oggetto chiamato "golden" ... una seconda ditta sta cercando di comprare delle "mele..."
- Ragionamento/interazione sulla base di ontologie differenti
 - Ontology mediators
- Estensione degli algoritmi di ricerca, con le ontologie e metodi di ragionamento

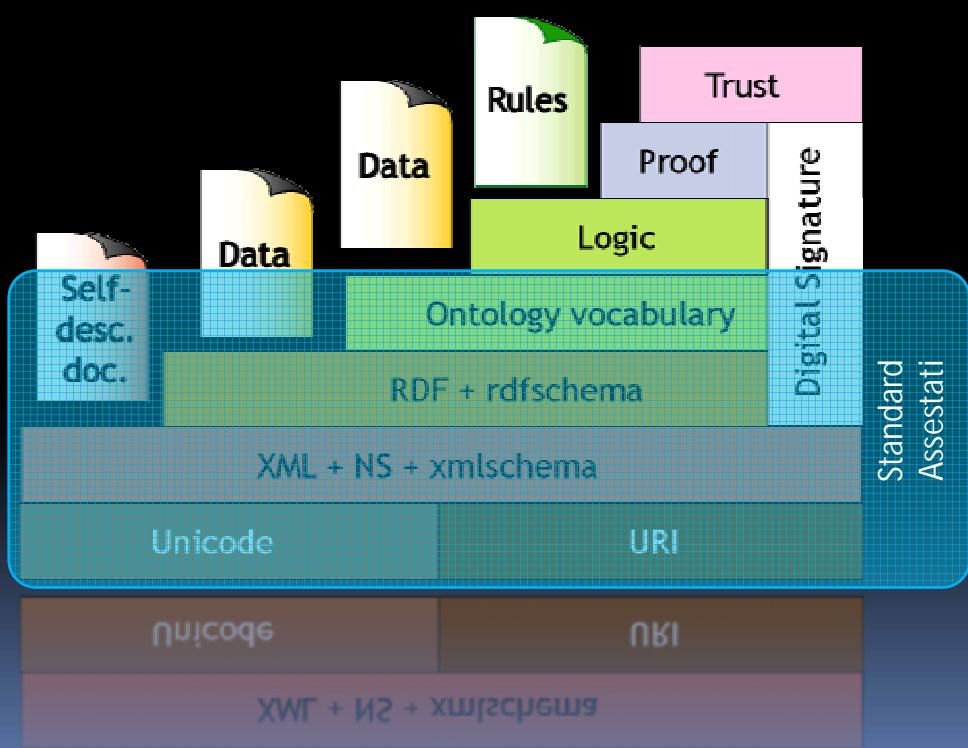
Esempi di ontologie

- Dublin Core, per documenti
- WordNet, sistema lessicale
- Gene Ontology, genomica
- Protein Ontology, proteomica
- SnoMed, campo medico (costo 1500\$/anno)

Strumenti per definire ontologie:

- Protégé <http://protege.stanford.edu/>

E gli altri strati di SW?



Regole e Logiche

- Le Description Logics si stanno affermando come logiche di riferimento per SW
 - Si sviluppano a partire dagli anni '80
 - Tecnologia assentata
 - Decidibili (la logica dei predicati è semi-decidibile)
 - Hanno complessità computazionale accettabile ($O(P)$)
 - Sono basate sul concetto di *ruolo* (non hanno le variabili)

Regole e Logiche

- Recentemente, un fortissimo interesse (industria e ricerca) per le *regole*
- Informazioni semantiche comprendono anche regole che descrivono aspetti caratteristici di un dominio
- Problema aperto: ragionare sia sui dati semantici che sulle regole
- **Rule Interchange Format**: standard W3C (tentativo) per la rappresentazione di regole
- Applicazioni a Semantic Web Services, Business Rules, e-Contracts,...

Proof e Trust ?

- Prevalentemente argomento di ricerca
 - Argumentation dialogues
 - Proof exchange
 - Trust models
 - Veridicità delle informazioni
 - Incompletezza delle informazioni
 - Informazioni contraddittorie

Semantic Web – problemi ?

- Nonostante il paradigma SW sia in discussione dal 2001 ...
- ... nel web attuale tale architettura stenta ad affermarsi
 - Molto discussa in ambito accademico
 - Poco apprezzata in ambito industriale
 - Solo recentemente, con i Semantic Web Services, è (ri)sorto un notevole interesse nell'ambito business
- In generale, “fatica” a percepire i vantaggi

Semantic Web – problemi?

- Adozione di RDF
 - Aggiungere contenuto semantico è molto costoso
 - Finché non si raggiunge una “massa critica” di informazioni semantiche disponibili su web ...
 - ... gli strumenti di SW appaiono essere “sterili”
 - Recente proposta in W3C di *Gleaning Resource Descriptions from Dialects of Languages* (GRDDL)
- Ontologie
 - Costo elevato nel produrre una ontologia ex-novo
 - Una ontologia è “viva”, muta nel tempo
 - Costo di manutenzione
 - Costo di gestione (uso di temporal logic per gestire il versioning)

Semantic Web – critiche?

- Praticamente irrealizzabile
 - Problemi intrinseci quali ad esempio il “metacrap”
- Quale uso dei dati?
 - Problemi di censura
 - Problemi di privacy
- I dati semantici sono già disponibili su web, è sufficiente estrarlarli opportunamente
 - SW inteso come una duplicazione non necessaria di ciò che esiste già
 - GRDDL potrebbe essere usato anche a tale scopo

Conclusioni

- Semantic Web: aggiunta di informazioni semantiche sul contenuto delle risorse presenti su web
- Grandi prospettive sulla carta ...
 - Fortissimo interesse nell'ambito della ricerca, specie nel settore dell'intelligenza artificiale
- ... fatica ad affermarsi in realtà
 - Grande interesse recentemente per i Semantic Web Services

Qualche riferimento...

Sito ufficiale W3C:

- <http://www.w3.org/2001/sw/>

Sito comunità

- <http://www.semanticweb.org/>
- <http://www.websemantico.org/>