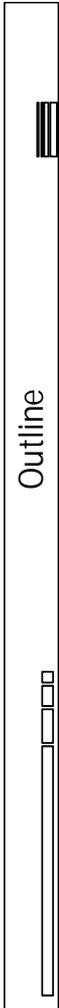




Ing. Federico Chesani

SEMANTIC WEB

12 Giugno 2008



Outline

- Introduzione a Semantic Web
- Tecnologie del Semantic Web
- SW oggi: problemi e critiche

Outline

Introduzione a Semantic Web

- Caratterizzazione del World Wide Web odierno
- La proposta Semantic Web
- Architettura del Semantic Web
- Applicazioni del Semantic Web

Il Web oggi (Web 1.0) ...

- L'informazione oggi è rappresentata tramite:
 - Linguaggio naturale
 - Immagini, multimedia, aspetto grafico
- Gli utenti utilizzano tutti questi mezzi (facilmente) per:
 - Dedurre fatti da informazioni parziali
 - Creare associazioni mentali
 - Accedono contemporaneamente a diversi mezzi di comunicazione senza alcun problema (uso di più sensi)

II 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

II Web oggi (Web 1.0) ...

sembrare tutto fatto, invece approvato un solo articolo. All'origine del blocco il no di Mastella all'art.91

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

Dinari ripete: "Mani libere fino all'ultimo". Predirettorista di CLAUDIA FUSANI



Lamberto Dini

ROMA – L' "assurdo" è andato in scena oggi nell'aula di palazzo Madama sfidando la logica, il buon senso, la correttezza gestionale della politica e lasciando il passo alle manovre di palazzo. Doveva essere il giorno dell'attesa e del fatidico voto finale alla finanziaria. Alle nove e mezzo del mattino, quando il presidente del Senato Franco Marini riprende i lavori mancano poche articoloni – dal 91 al 97 – e circa ottanta votazioni. Poche ore ed è finita. Alle otto di sera, invece, l'articolo 91 è stato accantonato – rinvitato – il 92 approvato, il 93 in piena discussione. E anche molto vivace. Si e no sono state fatte una decina di votazioni. Un'imposta assurda, kafkiana, che racconta una delle pagine più incredibili di questa

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

... se domani il designer 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>
```

II 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?
 - E.g.: Nella mia home page esistono link ad altre home page...
 - quali di questi sono i miei colleghi?
 - quali di questi sono semplicemente miei conoscenti?

II 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
 - Esempio: ultime elezioni americane

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

“The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.”

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
 - ... dei contenuti
 - ... delle informazioni di routing (links)

Aggiungere informazioni sul contenuto

Aggiungere informazioni sul contenuto non è sufficiente

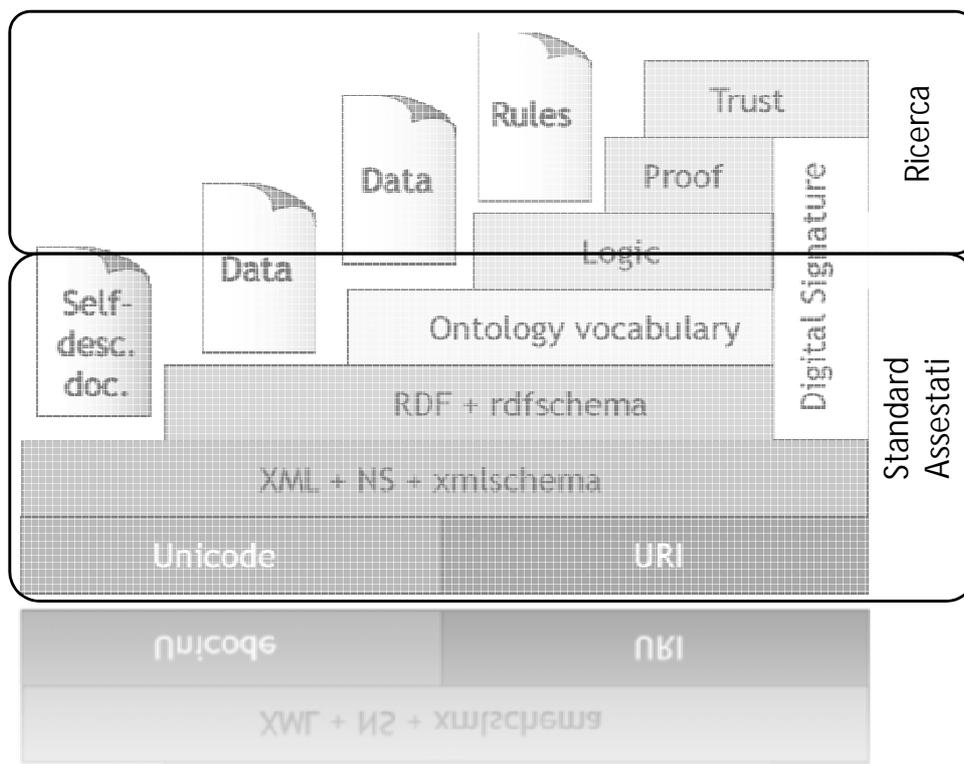
- Tali informazioni devono essere opportunamente *strutturate* (e.g., classificazione di Linneo per il mondo animale)
 - *Ontologie?*
- E' necessario aggiungere anche *regole di inferenza* (e.g., sillogismo aristotelico, FOL, DL)
 - *Logica?*
- Applicando regole a dati si ottengono nuove conoscenze
 - 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, contraddittorie ed incomplete

Architettura del Semantic Web



SW – Appl i cazioni ?

Qualche esempio ...

- Sistemi di ricerca
- Assistente remoto
- Integrazione Database (ambito industriale)
- Biblioteche ed archivi digitali (XMP Adobe)
- Semantic Web Services

Appl i cazioni Ricerca di documenti / risorse

- Aziende medio/grandi hanno il problema di gestire lo storico dei documenti interni
 - E.g., Google offre "(Mini) Search Appliance" (\$1,995 up to 50.000 docs, \$30,000 up to 500.000 docs – 2007 prices)
- Vodafone Live! Mobile Portal ha adottato una soluzione basata su RDF (SW)
 - Pagine visitate per download diminuite del 50%
 - In due mesi, scaricate il 20% in più di suonerie (fonte: Ivan Herman, SW lead)

Applicazioni Altri portali ...

- Sun's White Paper e System Handbook
- Nokia's S60 – portale di supporto
- Harper's Online magazine - collega gli articoli pubblicati tramite una ontologia interna
- Oracle - virtual press room
- Opera's community site
- Yahoo! Food
- FAO's Food
- Nutrition and Agriculture Journal - portale

Applicazioni Assistente Remoto

- Scenario originale proposto da Tim Berners-Lee
- Due utenti, tramite il loro "Semantic Web Agent" residente in un palmare/cellulare, riescono a:
 - Sincronizzare le rispettive agende
 - Generare un piano di azioni e su questo giungere ad un agreement
 - Delegare, in base ad un modello di trust, compiti l'uno all'altro
 - Analizzare clausole di un contratto ed effettuare scelte in base a funzioni di costo/benefici

Applicazioni Assistente Remoto rivisto

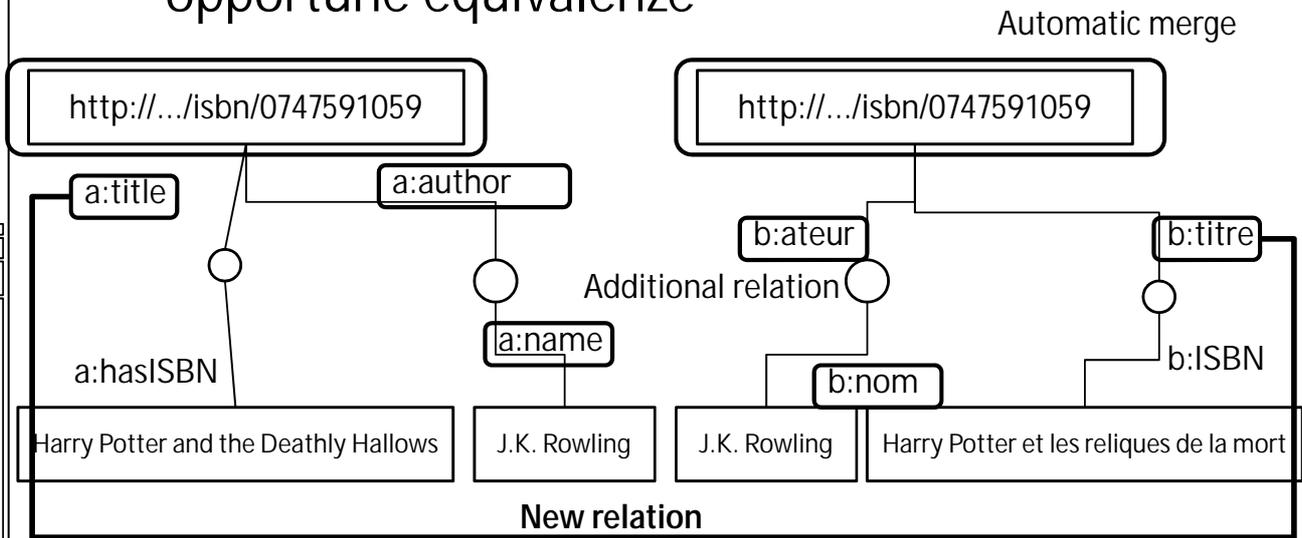
- Ognuno di noi può accedere *on-line* a:
 - c/c bancario
 - agenda
 - album fotografico
- Tali fonti di dati non sono integrati:
 - Non possiamo ad esempio integrare l'agenda con i movimenti bancari effettuati quel giorno...
 - Non possiamo collegare alcune fotografie al giorno in cui sono state scattate...
- Minor enfasi sulla parte di ragionamento!

Applicazioni Integrazione di database

- A livello intra-aziendale, è frequente la necessità di dover integrare differenti database in un unico nuovo modello
- In maniera approssimata:
 1. Si rappresentano i diversi dati tramite strutture astratte di rappresentazione
 - Dati indipendenti dalla loro rappresentazione interna
 2. Si fa il "merge" di tali rappresentazioni astratte
 - Merge *reale* tramite esportazione/importazione verso un nuovo db
 - Merge *virtuale*, tramite opportuni strati software che nascondono la realtà fisica
 3. Si utilizza il risultato ponendo query di complessità/significato maggiore (rispetto a prima)

Applicazioni di integrazione di database

- Il passo di merge ha successo se per ogni set di dati vengono identificate/definite opportune equivalenze



Applicazioni di integrazione di database

- Semantic Web offre supporto a tale lavoro di definizione di relazione tra concetti
- Differenze:
 - Prospettiva sull'intero web
 - De-centralizzazione dei dati
 - Uso di ontologie per organizzare meglio le relazioni
- Strumenti:
 - RDF
 - GRDDL

Applicazioni

Biblioteche ed archivi digitali

- La maggior parte dei sistemi di catalogazione di documenti avviene per categorie
 - E.g., è possibile consultare lo schedario di una biblioteca per *argomenti*
- Ciò non è possibile con il web odierno
- E.g.: se cerchiamo qualcosa attinente al concetto di “Intelligenza Artificiale” ...
- ...ci viene restituito un elenco di pagine che contengono la keyword “Intelligenza Artificiale” ...
- ... non è detto che le pagine siano effettivamente relative all'AI

Applicazioni

Semantic Web Services

“Un WS è un’applicazione software identificata da un URI, le cui interfacce pubbliche e relativi binding sono definiti e descritti in XML. La sua definizione può essere trovata da altre applicazioni software.

Questi ultimi possono poi interagire con il WS seguendo le direttive presenti nella definizione del servizio, usando messaggi XML trasportati da protocolli internet.”

(“Web Services Architecture”

<http://www.w3c.org/TR/2002/WD-ws-arch-20021114>)

Applicazioni Semantici Web Services

- Tramite UDDI e WSDL si possono conoscere dinamicamente (a tempo di esecuzione) locazione ed interfaccia di un servizio
- Però nessuna informazione su cosa faccia veramente il servizio in questione
- Es: posso invocare un servizio di nome `somma`, che riceve come parametri di ingresso due interi, e restituisce un intero. Chi mi assicura che il servizio svolga effettivamente la somma di due numeri?

Applicazioni Semantici Web Services

Sono necessari due tipi di informazione:

- Dati semantici su cosa faccia il metodo
 - E.g.: definito in termini di precondizioni, input, output ed effetti
- Regole su:
 - Come invocare il metodo
 - Quali vincoli ci sono sui dati

Applicazioni Semantic Web Services

Fortissimo interesse a livello sia di ricerca che di industria IT

- Due “contest” presenti in ambito di ricerca:

- IEEE Contest

<http://insel.flp.cs.tu-berlin.de/wsc06/>

- SWS Challenge

http://sws-challenge.org/wiki/index.php/Main_Page

Sponsorizzato da Stanford, vi partecipa da qualche mese anche IBM

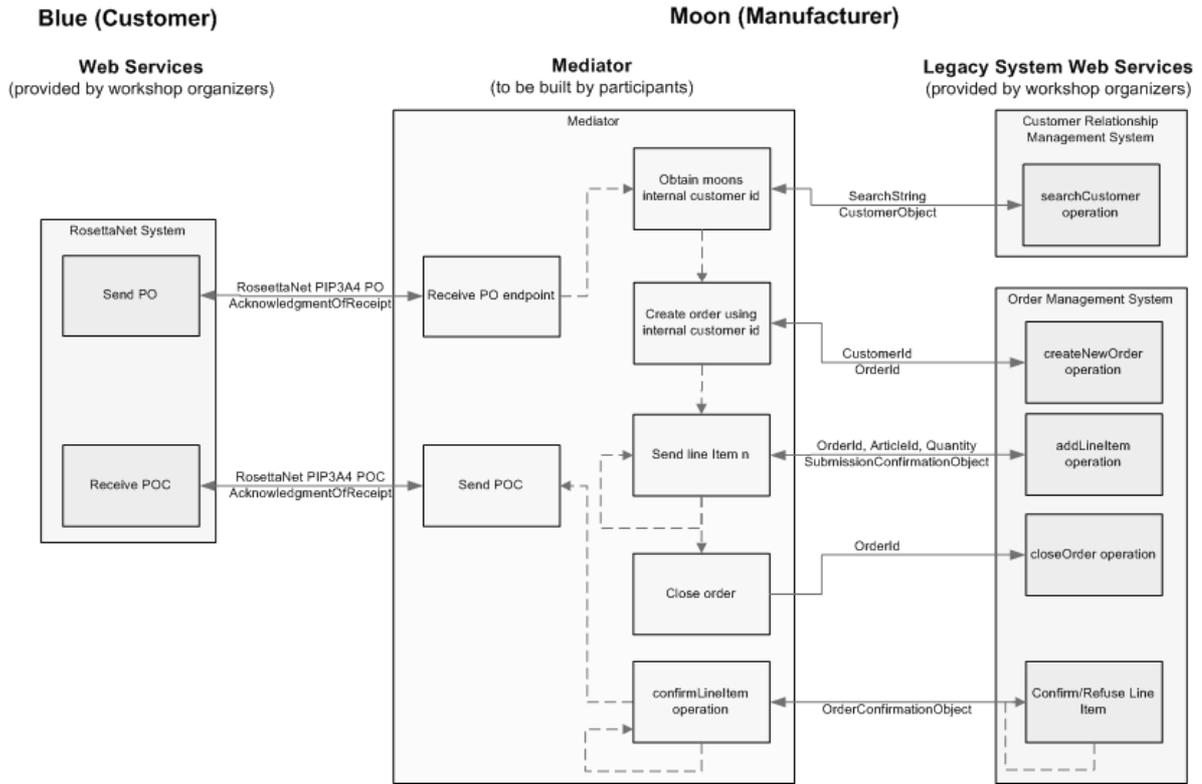
Applicazioni Semantic Web Services

SWS Challenge suddiviso in tre task differenti:

1. Date due società, Blue (customer) e Moon (manufacturer), costruire un mediatore per permettere l'interazione

- Blue utilizza RosettaNet PIP3A4
- Moon utilizza un legacy system interno
- I servizi dei singoli sono descritti tramite:
 - Informazioni sintattiche (WSDL e XML Schemas)
 - Descrizioni in linguaggio naturale

Applicazioni Semantici Web Services



Applicazioni Semantici Web Services

To: Smithers (Bristol)
No of packages: 1
Package dimensions: (l/w/h) 10/2/3 (inch)
Package weight: 20 lbs
For less than 120\$

2. Discovery di un servizio di spedizione, in base ad un certo obiettivo

- Sono fornite le descrizioni di 5 servizi di spedizione
- Si devono rappresentare le descrizioni, al fine

To: Szyslak (Tunis)
Package dimensions: (l/w/h) 40/10/10 (inch)
Package weight: 30 lbs
Current Time is 9:30 am
Shipping is required in 2 business days

- Discovery in base alla destinazione
- Discovery in base alla destinazione e al peso del pacco
- Discovery in base a destinazione, peso e prezzo
- Discovery con composizione di differenti servizi
- Discovery con deadline temporali

Applicazioni Semantic Web Services

3. Discovery (II) e composizione di servizi di vendita pc

- Descrizione in linguaggio fornitori
- Si deve supportare la
 - In base ad una definizione
 - In base a criteri più o meno
 - Componendo più servizi items

Apple Mac Book 13"
 Intel Duo Core Processor 2.0 GHz
 at least 512 MB RAM
 at least 60 GB HDD
 color: white or black
 at most 1500 \$

Apple Mac Book 13"
 at least Intel Duo Core Processor 1.8 GHz
 at least 512 MB RAM
 at least 60 GB HDD
 color: white or black
 at most 1500 \$

The price is most important to me, get me the cheapest offer if the other requirements are met

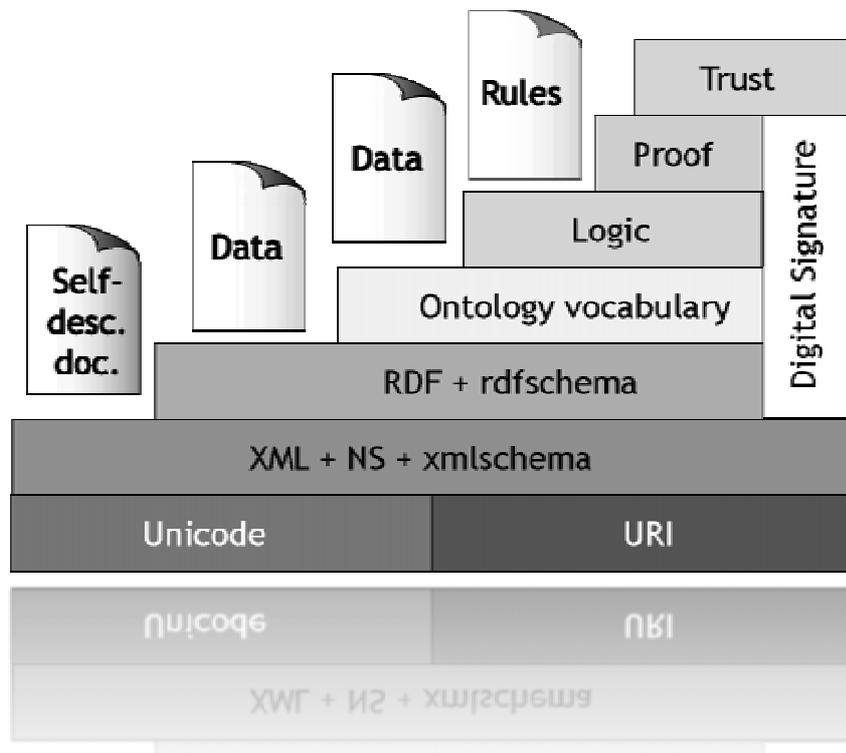
If the white notebook is significantly less expensive (more than 100\$), then buy the white one instead of the black one.

Outline

Tecnologie per il Semantic Web

- URI ed XML
- Informazione semantica (I): RDF e RDFS
- Informazione semantica (II): OWL
- Logiche, proof e trust

Architettura del Semantic Web



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!
 - non è necessario che vi corrisponda un contenuto

Esempi di URI:

<http://www.repubblica.it>
federico.chesani@unibo.it
 ISBN 88-7750-483-8

Uniform Resource Identifiers (URI)

- Gli URI sono identificatori unici usati nell'ambito di Internet
- Inizialmente (primi anni '90) partizionati in
 - Uniform Resource Locator (URL)
 - Uniform Resource Name (URN)
- Oggi tale distinzione non è più in "voga", ed URI ed URL vengono usati in maniera quasi equivalente

Esempi di URI:

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

Uniform Resource Locator

- Gli URL sono un tipo particolare di URI
- Identificano la risorsa tramite il principale meccanismo di accesso

`http://lia.deis.unibo.it/~fc/LIAIndex.html`

Schema di accesso

Network location

- Non è necessario che ad un URI/URL corrisponda per forza qualche "cosa"

eXtensible Markup Language - XML

- Nasce per l'interscambio di dati tra sistemi diversi
 - No informazioni di presentazione
 - Accessibile da un utente umano e da una macchina
- Estensibile a piacimento, al fine di meglio rappresentare i dati in questione
- Strutturato gerarchicamente tramite *tags*
- Un documento XML può contenere una descrizione della sua grammatica (opzionale)

- Tecnologia molto assestata ed oramai ampiamente utilizzata

eXtensible Markup Language - XML

- Offre due linguaggi ulteriori per specificare la grammatica di un dialetto XML
 - Document Type Definition (DTD)
 - XML Schema
 - Tipi di dato
 - Namespaces
 - Usa sintassi XML a sua volta
 - Maggiore potere espressivo nel vincolare il contenuto e la strutturazione dei nodi figli
- Concetto di documento valido e documento ben formato

eXtensible Markup Language - XML

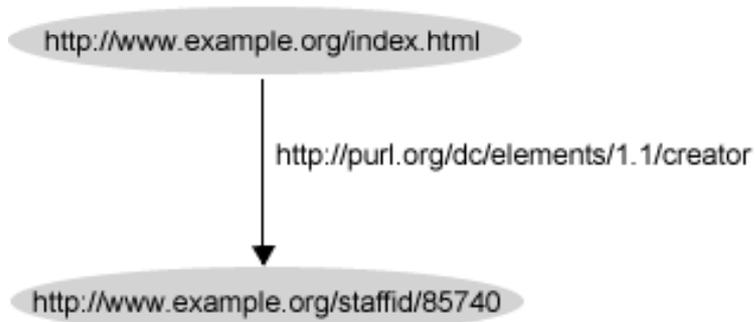
- Esistono framework completi per la gestione di documenti XML (DOM e SAX)
- Linguaggio XSL (XML Stylesheet language)
 - XSLT (XSL Transformation)
 - XPath, linguaggio per espressioni (query/matching su documenti XML)
 - FO (Formatting Objects)

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, predicato, oggetto >
 - < risorsa, attributo, valore >
- Diverse rappresentazioni possibili (N3, Grafo, RDF/XML)

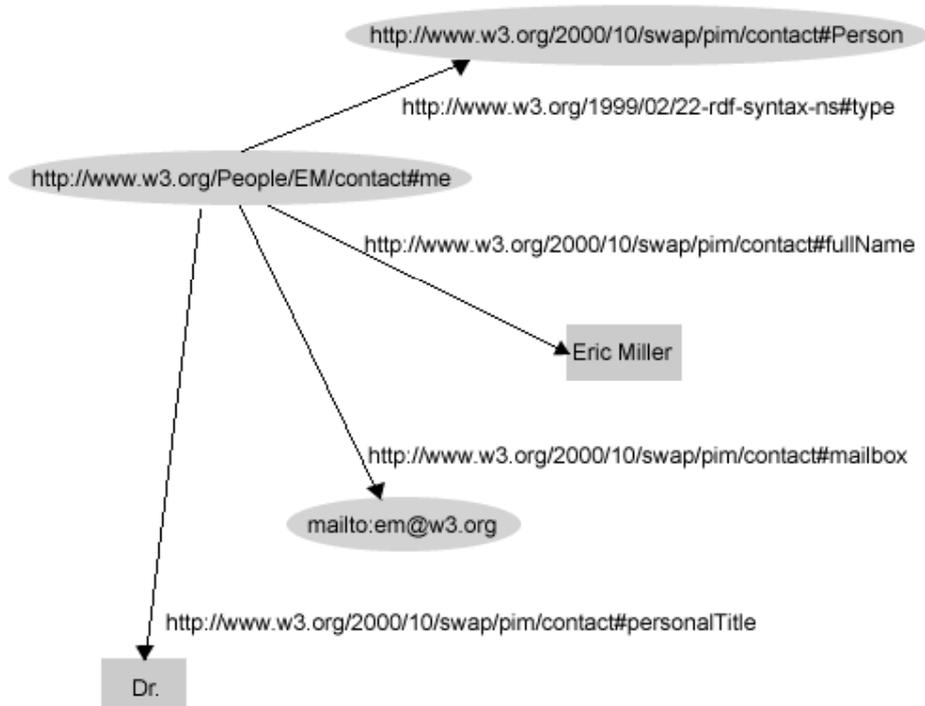
RDF – Rappresentazione a grafo

- 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

xml ns: rdf=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>

xml ns: 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" />

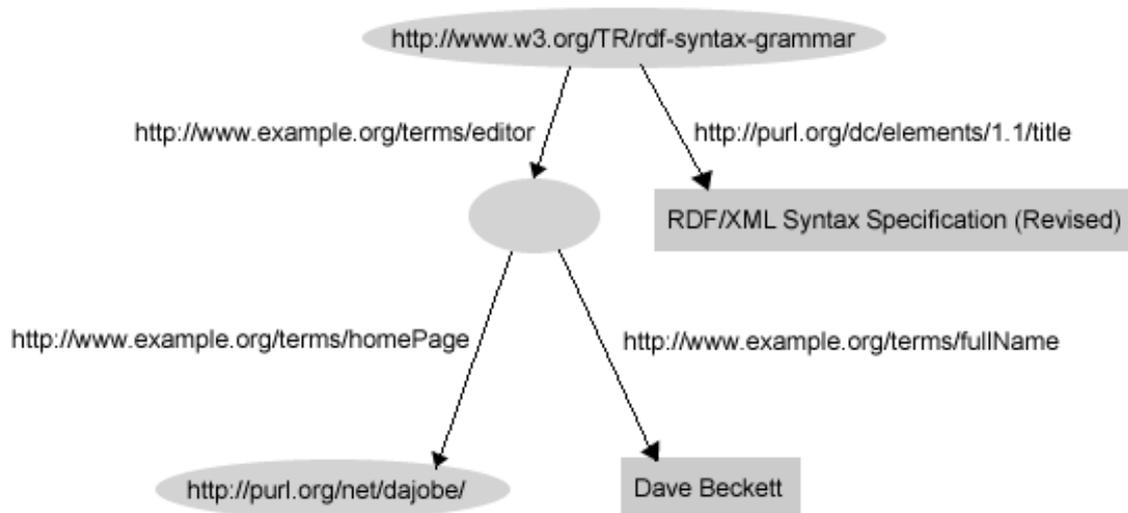
 <contact: personalTitle>Dr. </contact: personalTitle>

</contact: Person>

</rdf: RDF>

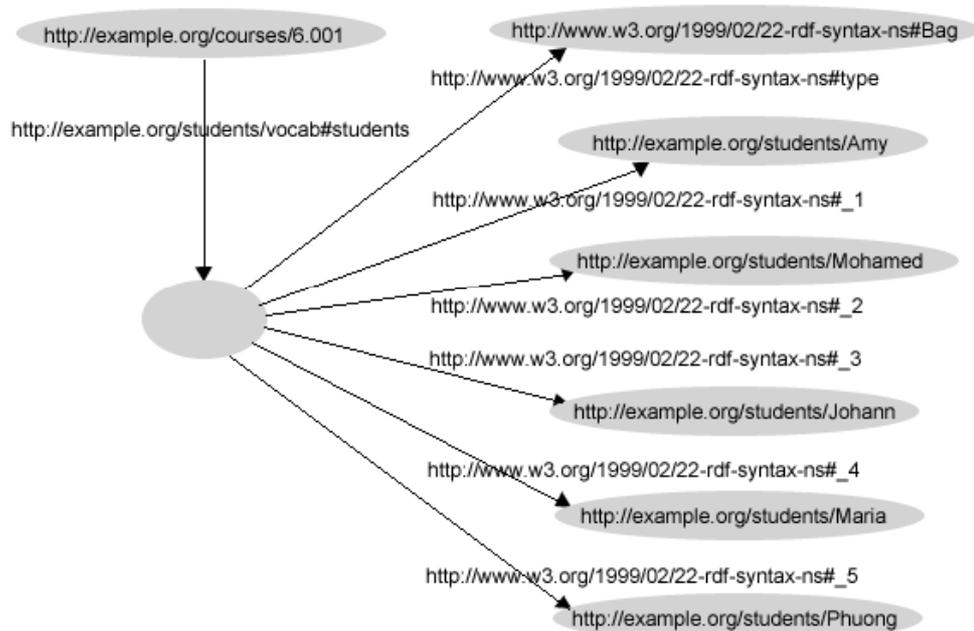
Posso chiedere al computer la mailbox di Eric Miller, senza sapere a priori se usa una casella di posta elettronica, piuttosto che una casella postale fisica, o quant'altro...
... e se Eric Miller cambia mailbox, il risultato della mia ricerca sarà sempre coerente!

RDF – esempi



Empty Nodes

RDF - esempi



Bags/Sets

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, alternative)
- *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 Modelli E/R

- Molte similarità con il modello E/R...
 - ... ma rdf risulta essere più espressivo
- RDF inteso come E/R per il web
- Le relazioni in RDF sono “first class entities”
- In RDF l’elenco delle proprietà di una entità non è:
 - Fissato a priori
 - Centralizzato in un unico punto (DB)
 - Conseguenza del fatto che chiunque può dire qualunque cosa di chiunque altro

RDF e Database Relazionali

Esiste un mapping diretto con i DB relazionali

- Un record è inteso come un nodo RDF
- Il nome di una colonna è inteso come `rdf:propertyType`
- Il corrispondente campo nel record è inteso come valore
- RDF è pensato per integrare più database con modelli differenti
 - DBMS tradizionali sono invece ottimizzati per costruire nuovi modelli di dati all'interno dello stesso DB o nell'ambito di un numero ristretto di altri DB

Framework per RDF

- JENA, framework Java per la rappresentazione di RDF statements
- SPARQL, Linguaggio di query per Repository RDF, che supporta la navigazione dei grafi anche su Repository diversi e/o distribuiti
 - Agnostico rispetto al linguaggio di implementazione

Strumenti per RDF

Moltissimi strumenti già disponibili...

Solo nella wiki di W3C sono elencati:

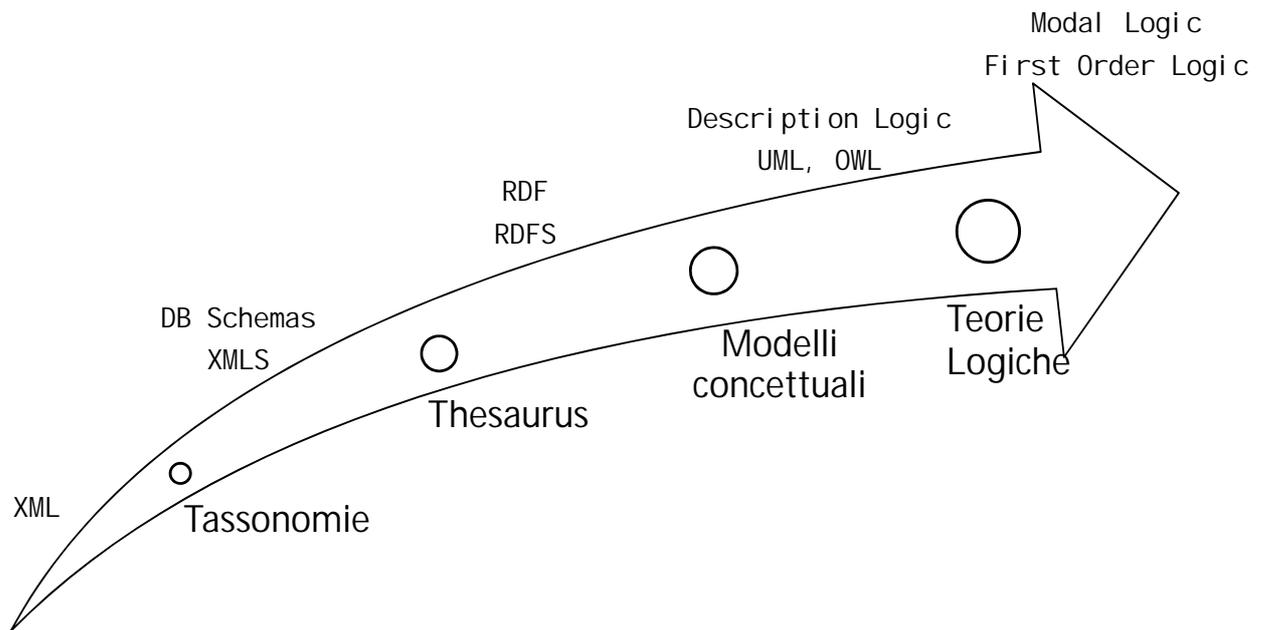
- Editors e framework per più di 14 linguaggi
 - Adobe XMP
 - IBM Semantic Layered Research Platform
- Repository RDF: 24 sistemi
 - Oracle Spatial 10g

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, omonimia, meronimia)
 - risolvono problemi tipici del linguaggio naturale, quali ambiguità e risonanza
 - Esempio: WordNet

Modelli semantici

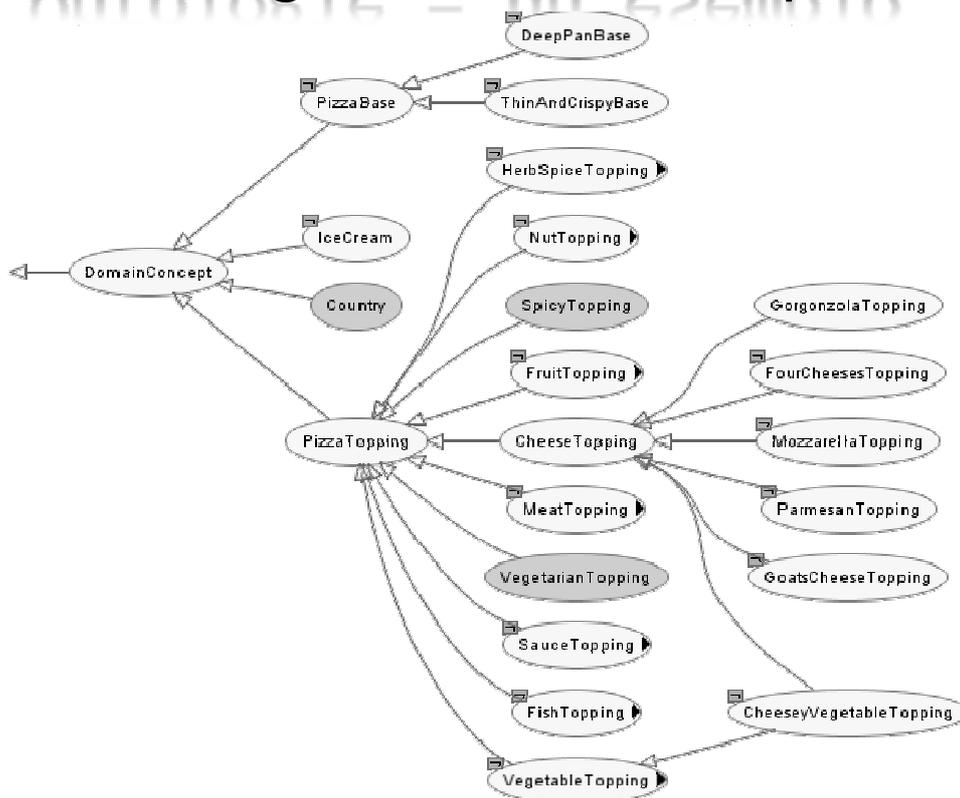
- Modello concettuale: è il modello di una particolare area di conoscenza (*dominio*).
Definisce:
 - Entità del dominio
 - Relazioni fra le entità (in forma di proprietà ed attributi)
 - Regole su classi, attributi e relazioni
 - Regole di inferenza ... *Teorie Logiche!!!*

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



Ontologie

Necessarie a prescindere da Semantic Web

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

- Il problema risiede nelle “sottili distinzioni di significato”
- Oggetto di studio in AI fin dagli anni '80, per supportare *knowledge sharing/integration*

Ontologie

Chi ne ha sentito la mancanza?

Fallimento delle dotCom (2001)

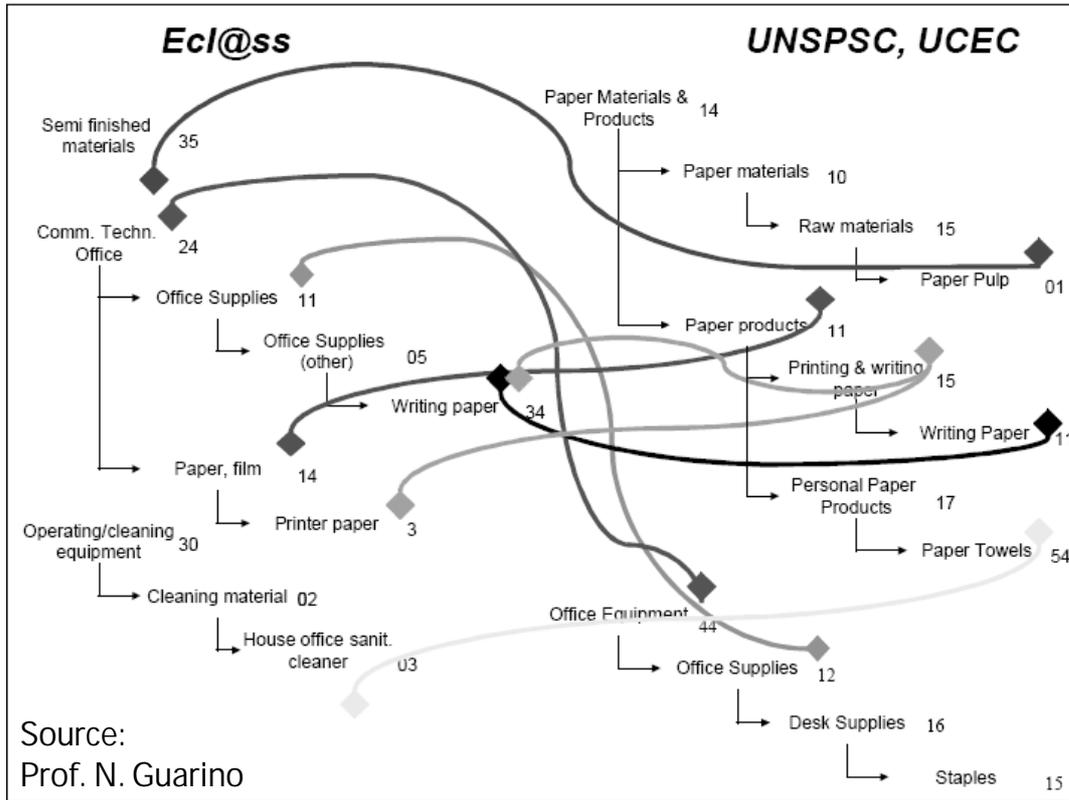
Harvard Business Review, October 2001:

“Trying to engage with too many partners too fast is one of the main reasons that ***so many online market makers have founded.***

The transactions they had viewed as simple and routine actually involved many ***subtle distinctions in terminology and meaning***”

Ontologie

Chi ne ha sentito la mancanza?



Source:
Prof. N. Guarino

Ontologie

Non basta XML?

“XML is only the first step to ensuring that computers can communicate freely. ***XML is an alphabet for computers*** and as everyone who travels in Europe knows, knowing the alphabet doesn't mean you can speak Italian or French”

Business Week, March 18, 2002

Interesse per le ontologie

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

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, non decidibile)

OWL – caratteristiche

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

Problemi ontologici e sol.

- Equivalenza tra ontologie diverse
 - Ontology matching problem – alignment 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

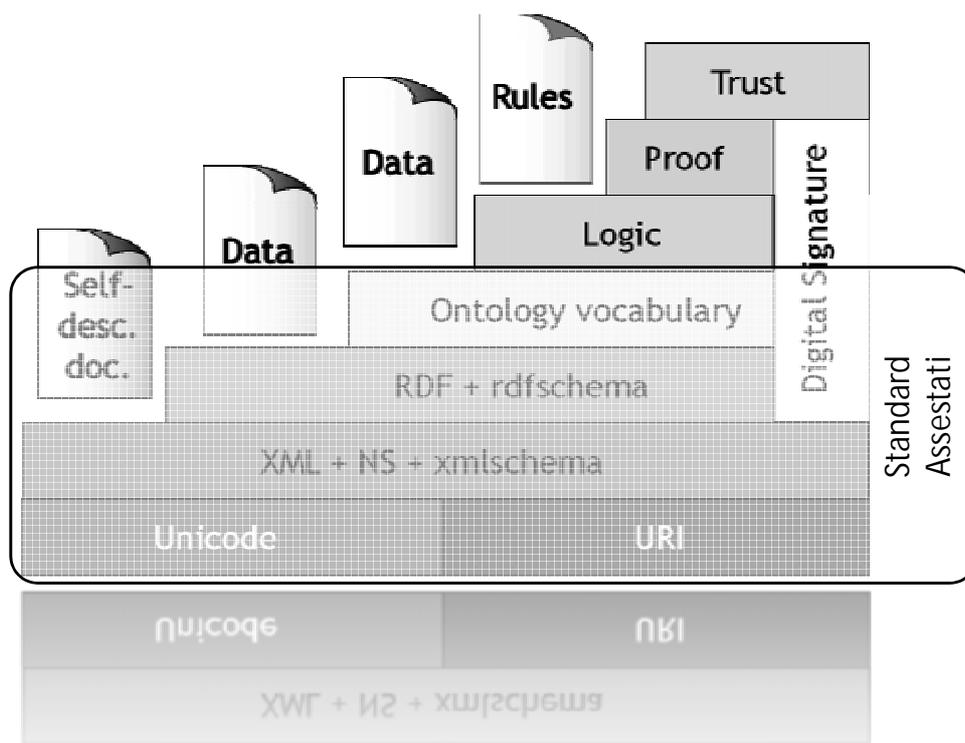
Esempi di "ontologie"

- 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?



Logiche

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

Description Logics

- Progettata per la descrizione delle definizioni e delle proprietà delle categorie (dei concetti)
- Concetti
 - Atomici (e.g. uomo)
 - Complessi (e.g. (and persona maschio))
- Ruoli (e.g., (:età 27))
- Costanti (e.g., johnSmith, federicoChesani)

Le costanti sono per i singoli individui, i concetti per classi di individui, e i ruoli per relazioni binarie (nei frames, si parla di *individual frames*, *generic frames*, e *slots*)

Descrizione Logica - Sintassi

Simboli non logici:

- Concetti atomici
- Ruoli
- Costanti

Simboli logici:

- punteggiatura ed interi positivi, ...
- operatori di costruzione dei concetti (ALL, EXISTS, FILLS, AND, e moltissimi altri!!!)
- connettivi ($=_{\text{def}}$, $:<$, \rightarrow)

N.B.: Attenzione alla notazione, che non è univoca!!!

Descrizione Logica

Dato un dominio finito e ground, ogni concetto determina una *estensione*, cioè l'insieme di tutti gli individui del dominio cui il termine si applica

- Definizione di un individuo:

```
federico =def (  
  :natIn italia  
  :età 33  
)
```

- Definizione di un concetto:

```
cittadino_italiano =def (  
  (and  
    persona  
    (ALL :natIn italia)  
  )  
)
```

Description Logic

Operatori di costruzione dei concetti:

- (EXISTS n r): tutti gli individui che sono in relazione r con almeno altri n individui
- (FILLS r c): tutti gli individui che sono in relazione r con l'individuo c
- (ALL r d): tutti gli individui che sono in relazione r *solo* con individui descritti da d
- (AND d1...dn): tutti gli individui descritti da tutti i concetti d1...dn (intersezione)
- (at-most n r): tutti gli individui che sono in relazione r con al massimo n individui
- (one-of i1...in): uno degli individui tra i1...in
- ...

Description Logic

"Courses taken by 60 to 90 students, who are all undergrads, and taught by a CS professor"

(and
 COURSE
 (at-least 60 takers)
 (at-most 90 takers)
 (all takers (and STUDENT
 (all inYear (one-of 1 2 3 4))))
 (exactly 1 taughtBy)
 (all taughtBy (and PROFESSOR
 (fills inDepartment "CS"))
)
)
)

Description Logic

“Frase” del linguaggio:

uomo = _{def} (and persona maschio)	<i>definizione</i>
uomo :< persona	<i>definizione parziale</i>
joe → (and uomo chirurgo)	<i>proprietà di un individuo</i>

La conoscenza è specificata tramite Abox (assertion box, riguardo i singoli individui) and Tbox (terminological box, riguardo i concetti, le relazioni tra di essi, i ruoli, etc.)

Description Logic

Principali metodi inferenziali

- Sussunzione
 - Una categoria è il sottoinsieme di un'altra
- Classificazione
 - Un oggetto appartiene ad una categoria?
- Consistenza di definizione di categoria
 - La definizione è logicamente insoddisfacibile?

Descrizione Logica

Sussunzione

- Data la definizione terminologica di gatto
- Data la def. term. di animale_domestico
 - Non è detto che vi sia una relazione diretta tra i due termini
- Dimostrare che gatto è sussunto (cioè è un iponimo, una specializzazione) di animale_domestico

gatto :< animale_dom

Regole

- 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,...

Regole – di diversi ambiti applicativi

Negli ultimi 3 anni è sorto un notevole interesse in ambito scientifico per l'uso delle regole...

...interesse motivato da alcuni interessi economici di tipo industriale

- Uso di regole per la definizione di processi
 - Controllo, verifica a priori e a run-time, enactment di processi
 - Business rules e ricerca di agreement
 - E-Contracts
 - Choreografie, Interoperability, paradigma "off-the-shelf"

Strumenti per SWS

- Diversi standard proposti, nessun vincitore
 - Semantic Annotations WSDL (SAWSDL) (W3C)
 - Semantic Web Service Language (SWSL) (W3C)
 - OWL Web Services Ontology (OWL-S)
 - Supportata negli US da DAML (DARPA Agent Markup Language)
 - Web Service Modeling Ontology (WSMO)
 - Supportata principalmente in Europa dal DERI (Digital Enterprise Research Institute: Galway, Seoul, Stanford)

Proof e Trust ?

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

Outline

Conclusioni

- Problemi nella diffusione di SW
- Critiche a SW
- Considerazioni finali
- Qualche link

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
 - Meno apprezzata in ambito industriale
 - Solo recentemente, con i Semantic Web Services, è (ri)sorto un notevole interesse nell'ambito business
- In generale, "fatica" a percepirla 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 estrapolarli opportunamente
 - SW inteso come una duplicazione non necessaria di ciò che esiste già
 - GRDDL potrebbe essere usato anche a tale scopo

Semantic Web – critiche?

- Computazionalmente difficile
 - ... ma la Description Logic è pensata proprio tenendo in mente tale dimensione
 - ... non tutte le applicazioni necessitano di espressioni di arbitraria complessità (la complessità computazionale degli algoritmi si riduce, fino ad $O(P)$)

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

Conclusioni

- Recentemente, c'è chi propone di sostituire Semantic Web con Data Web
- Enfasi non più sulle possibilità di ragionamento...
- ...quanto sul fatto che alla fine si tratta di esporre in maniera organica, strutturata e comprensibile (standard) informazioni già disponibili

Considerazioni personali

- Chi dovrebbe inserire informazioni semantiche sui contenuti web?
 - Singoli utenti a titolo personale
 - Fenomeno del metacrap
 - Fenomeno delle folksonomies (e.g. flickr)
 - Fenomeno wiki
 - Aziende
 - Possiedono già database con semantica più o meno definita
 - Scenario di mercato globale estremamente concorrenziale

Considerazioni personali

- Esistono fattori culturali importanti nell'organizzazione delle aziende italiane
 - Non è detto che un accesso facilitato alle informazioni sia sempre desiderabile
- SW molto utile a livello *intra-aziendale*
- Alcuni dubbi sul livello *inter-aziendale*
 - *Business secrecy, nda, e altre pratiche commerciali possono essere un ostacolo*
- Moltissimi dubbi sulla "disclosure" a livello di web pubblico
 - Una azienda deve poter controllare minuziosamente
 - Quali dati rende pubblici
 - Che uso ne viene fatto (competitors?)

Qualche Link...

Sito ufficiale W3C:

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

Sito comunità

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

Grazie per l'attenzione

Questions?

Ing. Federico Chesani, Phd

c/o DEIS – Facoltà di Ingegneria

Viale Risorgimento 2

40136 - Bologna

federico.chesani@unibo.it

Tel. 051 20 93086