

Semantic Web

Underneath the new Wave

Luigi Iannone

Semantic Web: Cos'è?

- Nuova tendenza del Web
- Risposta ai problemi del Web
- Estensione Standard del Web
- Idea di (Sir) Berners-Lee (http://www.w3.org/2003/12/timbl_knighted)

Problemi del Web

- Information Overload
 - Sinonimi
 - Polisemi
 - Ambiguità in genere
- Informazione non machine-processabile
 - HTML (infos mischiate con grafica)
 - Nessun ragionamento

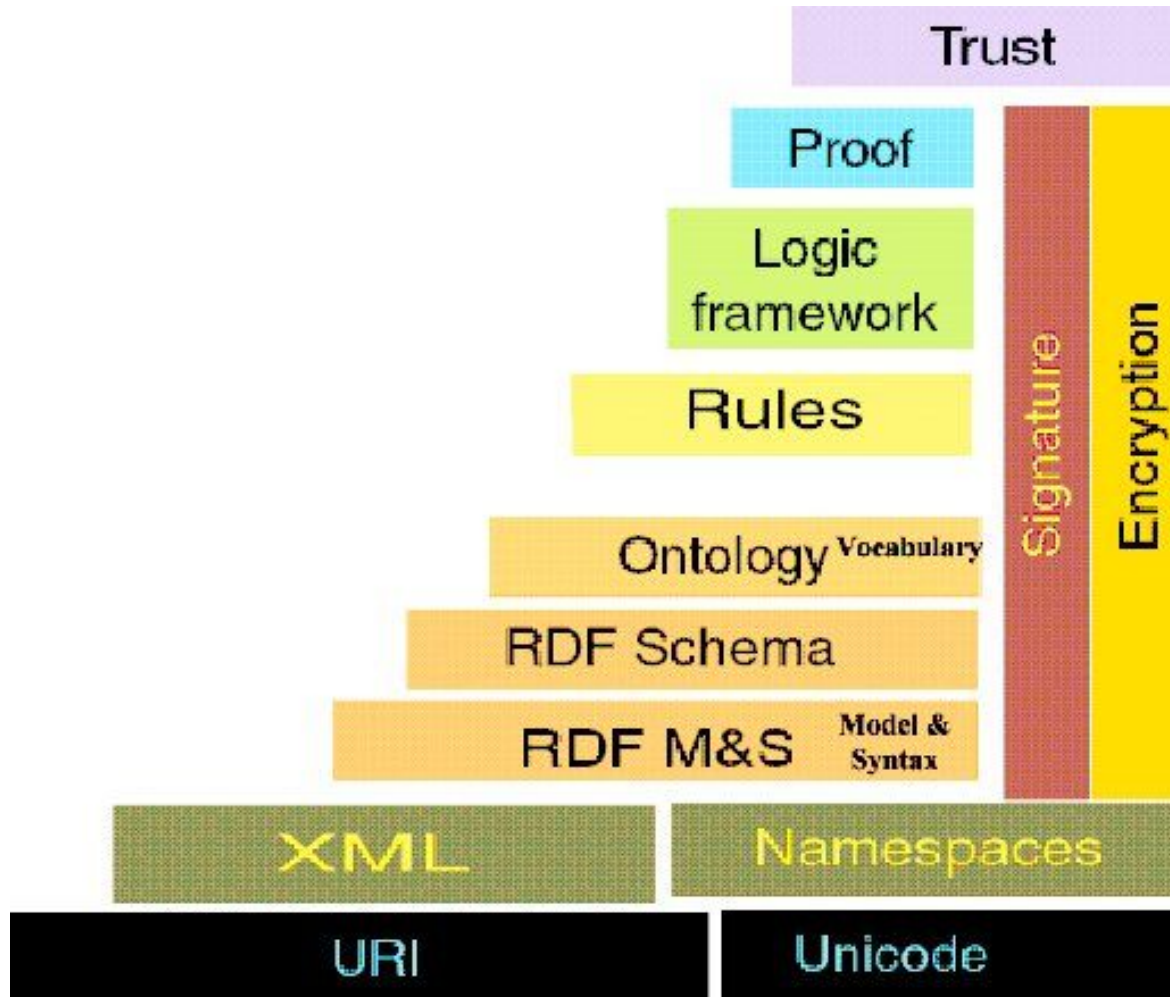
Risposte del Semantic Web

- Ontologie
- Metadati
- Logica
- Ragionamento



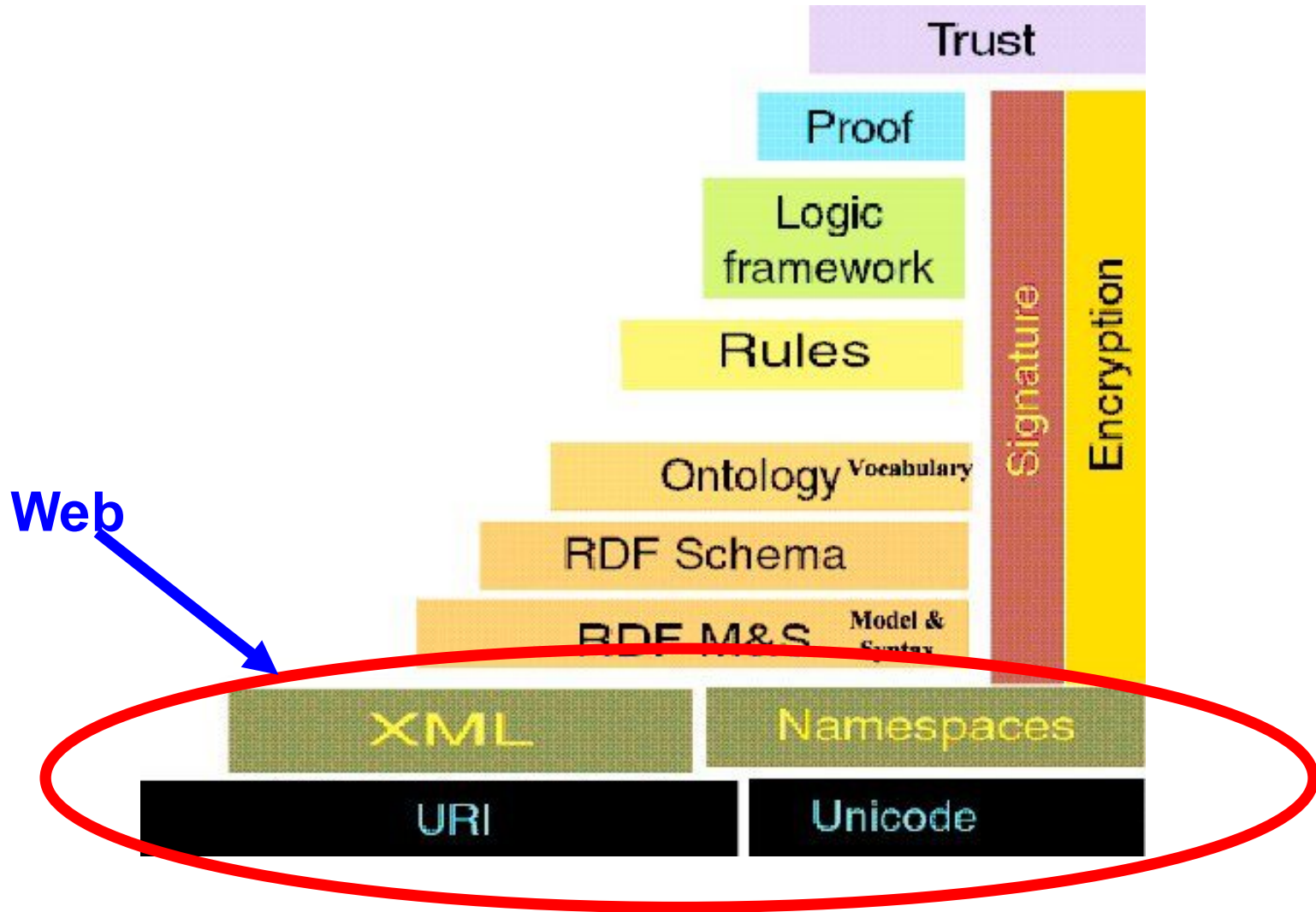
Informazione machine-processable

Semantic Web: Architettura



<http://www.w3.org/DesignIssues/diagrams/sw-stack-2002.png>

Semantic Web vs. Web



Riuso degli standard

- URL
- URI
- Unicode
- XML
- HTTP

Semantic Web: Livelli

- Architettura a Pila (analogo a Pila ISO/ OSI, Pila Protocolli Internet)
- Un livello → Un set di responsabilità
 - Information hiding
 - Alta coesione interna
 - Basso accoppiamento tra livelli



Architettura manutenibile e poco costosa

Livello sintattico

- Responsabilità
 - Concetto di documento univoco e astratto
 - Strutturazione gerarchica di documenti Web
- Tecnologie (standard)
 - URI/ URL
 - XML/ XML Schema/ DTD/ Namespaces
 - HTTP

Livello dei Metadati

- Responsabilità:
 - Concetto di risorsa (web e non web)
 - Annotazione standard delle risorse
- Tecnologie (standard)
 - RDF (serializzabile XML)
 - Tutto è una URI
 - Tutto va descritto con la massima semplicità possibile
 - Tutto può dire tutto su tutto

Livello Ontologico

- Responsabilità:
 - Vocabolari condivisi
 - Interoperabilità semantica (Le applicazioni *sanno* di cosa stanno parlando)
- Tecnologie (standard)
 - RDF Schema → DAML+ OIL → OWL
(serializzabili in XML)

Livello Logico

- Responsabilità:
 - Inferenza
 - Regole al top dell'ontologia
- Tecnologie (standard)
 - RuleML + OWL = OWLRules

Livello della prova e della fiducia

- Responsabilità:
 - Risoluzione dei conflitti
 - Web of Trust
- Tecnologie (standard)
 - Nessuno standard
 - Digital Signature
 - Catene di (s)fiducia, Modelli Friend of a Friend

Semantic Web: Scenari

- Ricerca di informazioni su Internet
- Classificazione di risorse
 - Item recommendation
 - e-commerce
 - ...
- Semantic Web Services

Ricerche su Internet (Esempio)

Voglio sapere chi è il direttore dell'
M.I.T.

Apro Google (<http://www.google.com>)

- Inserisco MIT Chair
- Risultati: 815.000 documenti che riguardano:
 - Massachussets Institute of Technology
 - mit (Tedesco) prep. *con*
 - Chair (Inglese) n. *sedia*
 - ...

Classificazione di risorse ...

Dato un insieme di articoli voglio selezionare il revisore più adatto conoscendone le competenze

Tecnologie Tradizionali

- Ogni articolo n keywords
- Ogni revisore m keywords
- Intersezione (poco probabile)

... Classificazione di risorse

...

Dato un insieme di articoli voglio
selezionare il revisore più adatto
conoscendone le competenze

Semantic Web

- Ogni articolo n concetti (di un'ontologia)
- Ogni autore m competenze (di un'ontologia)
- Intersezione ed inferenza (probabile)

... Classificazione di risorse

Dato un insieme di Items i.e. articoli di un qualsiasi (e-) shop ed un utente voglio *indovinare* quali articoli suggerire ad un generico utente

Approccio

- User Modeling
- Semantic Indexing

Semantic Web Services

- Evoluzione dei Web Services
- Applicazioni dinamicamente interfacciate
- Business 2 Business potenziato



Semantic web: Come?

Obiettivi

- Astrazione
 - Scambio indipendente da piattaforme e formalismi proprietari
- Disambiguazione
 - Sinonimia, Polisemia
- Ragionamento
 - Classificazione, Interoperabilità semantica

... Semantic Web: Come? ...

- Riutilizzo del Web:
 - Ri-scrittura dell'informazione



Vs.

- Arricchimento delle risorse correnti



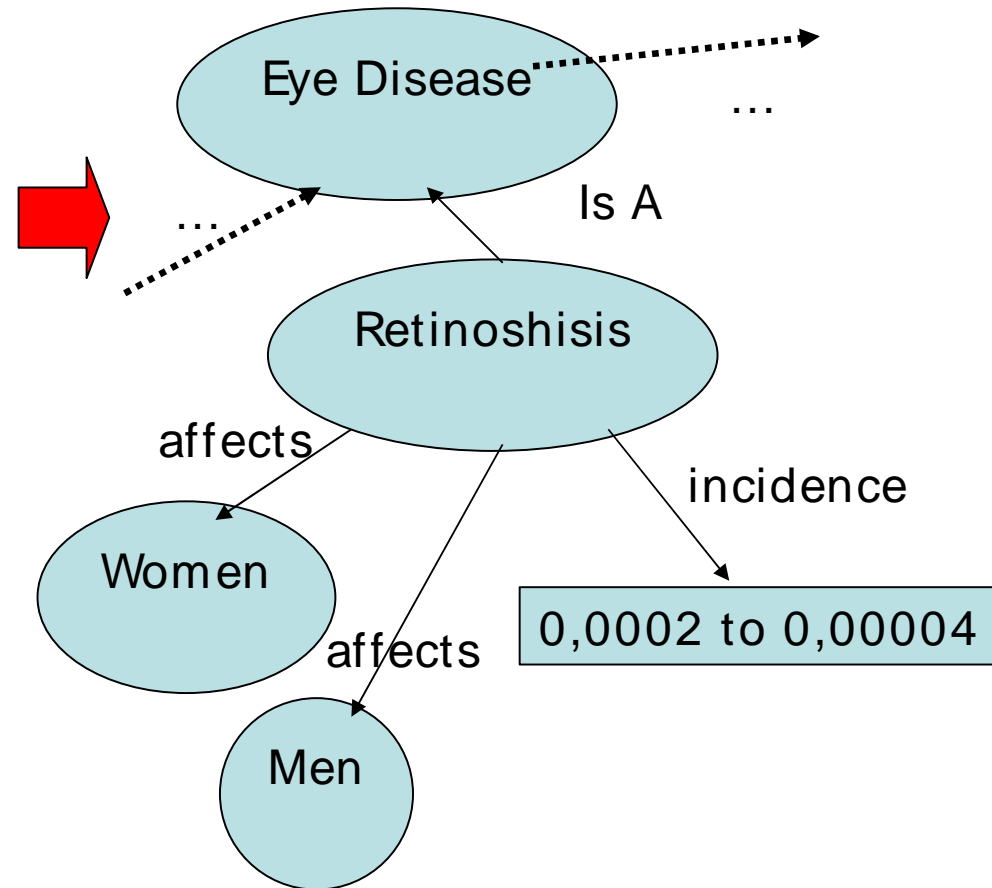
... Semantic Web: Come? ...

- Inserimento di Meta-informazione
 - Informazione sull'informazione
 - Formalismo Standard
 - Vocabolario condiviso
 - Ragionamento su metadati

... Semantic Web: Come? ...



Articolo su eMedicine
sulla retinoschisi
giovanile
<http://www.emedicine>



Ipotetico processo

- Estrazione di contenuto rilevante da testo non strutturato
- Trasformazione dell'informazione rilevante in un formato adatto alle fasi successive
- Prima fase di apprendimento automatico (addensamento dei contenuti) tipo clustering
- Estrazione di una teoria primordiale basata sull'output della fase precedente (prima concettualizzazione)
- Raffinamento della teoria (passo che può essere reiterato)
- Fusione con ontologie preesistenti

Estrazione di informazione rilevante

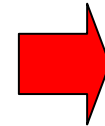
Obiettivo:

Individuare strutture nel testo che
possano essere frammenti di definizioni
di concetti.

... Estrazione di informazione rilevante...



Raw Test



Retinoschisis juvenile is an eye disease

Retinoschisis affects men

Retinoschisis affects women

Retinoschisis has incidence from 0,0002 to...

Concept examples description

... Estrazione di contenuto rilevante...

Problemi

- Di solito la forma dei testi, in qualsiasi lingua è molto meno lineare e strutturata di quella delle descrizioni soggetto predicato oggetto che vorremmo raggiungere
- Si potrebbe correre il rischio di perdere molta informazione (frasi coordinate, anafore...)

... Estrazione di informazione rilevante

Approccio iniziale

- Estrazione di Noun Phrases (NP): Frasi che contengono gruppi nominali
- Utilizzo di Shallow Parsers (SP)

Risultati:

- Estrazione di gruppi soggetto- verbo- oggetto (😊)
- SP attuali molto imprecisi (soprattutto con frasi lunghe (😞))

Trasformazione del contenuto rilevante

Obiettivo:

L'output di questo processo è ottenere un insieme di descrizioni in un formato standard adatto (e.g.: descrizioni RDF) risultanti dalla trasformazione delle NP ottenute nella fase precedente .

... Trasformazione del contenuto rilevante ...

Problemi

- In questa fase vanno gestite sinonimia e polisemia
- C'è bisogno di algoritmi di disambiguazione

... Trasformazione del contenuto rilevante

Approccio iniziale

- Trasformazione Noun Phrases+ verbo e oggetto in RDF
- Codifica dei nomi con le URI (ove presenti) dei synsets di WordNet

Risultati:

- Informazione strutturata su cui si può operare 😊
- Nessun algoritmo di word sense disambiguation provato 😞

Prima fase di apprendimento automatico (comincia lo spettacolo)

Obiettivo

Raggruppamento delle descrizioni precedentemente ottenute in gruppi omogenei. Senza alcuna supervisione (~ clustering)

... Prima fase di apprendimento automatico ...

Problemi

- Dipende dall'ordine della rappresentazione (su rappresentazioni di ordine ≥ 1 non vi sono risultati incoraggianti finora)
- Cluster omogenei di istanze

... Prima fase di apprendimento automatico ...

Approccio iniziale

- Descrizioni RDF di partenza come esempi
- Algoritmi di apprendimento automatico da esempi

Attendere prego...

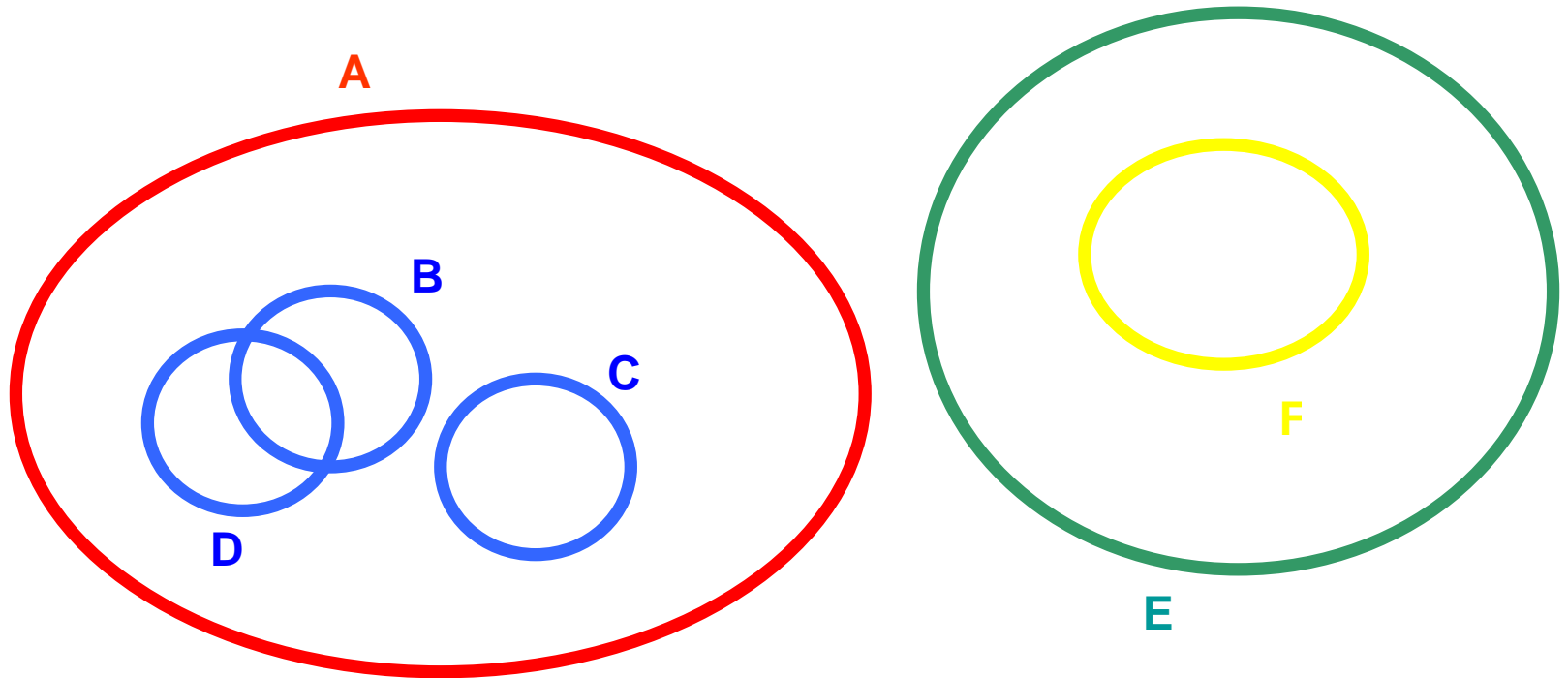
F



:

Attendere prego...

Estrazione di una teoria primordiale



$A \Leftrightarrow \dots$ $C \Leftrightarrow A \text{ and not}(B \text{ and } C) \dots$
 $B \Leftrightarrow A \text{ and } \dots$ $D \Leftrightarrow A \text{ and } \dots$

Raffinamento della teoria

Obiettivo:

Perfezionare la concettualizzazione
ottenuta per ottenere modelli più
predittivi

(Apprendimento supervisionato)

Apprendimento supervisionato

Input

- Definizione di un concetto (dalle fasi precedenti)
- Esempi positivi e negativi
- Definizione che copre qualche negativo o non copre qualche positivo (→ definizione erranea)

Output

Nuova definizione che minimizza l'errore

Fusione con ontologie preesistenti


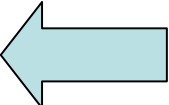
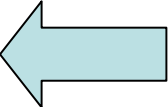
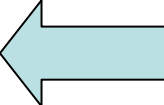

Obiettivo:

Fusione della
struttura con
quanto già
preesistente ed
integrazioni
eventuali in
entrambe le
direzioni



Still pure dream

Ricerca in corso...*

- Estrazione di contenuto rilevante da testo non strutturato
- Trasformazione dell'informazione rilevante in un formato adatto alle fasi successive 
- Prima fase di apprendimento automatico (addensamento dei contenuti) tipo clustering 
- Estrazione di una teoria primordiale basata sull'output della fase precedente (prima concettualizzazione) 
- Raffinamento della teoria (passo che può essere reiterato) 
- Fusione con ontologie preesistenti 

* Le dimensioni delle frecce non sono casuali

Stato dell'arte (Semantic Web)

- Metadati → Resource Description Framework (RDF)
- Concettualizzazioni Strutturate (Ontologie)
 - DAML+ OIL, OWL

Rappresentazione della conoscenza (KR)

- Logica del primo ordine (FOL) o suoi frammenti
 - Completamente esplorata
 - Flessibile
 - Direttamente integrabile con i DB
- Logiche Descrittive (DL, DLs)
 - Conoscenza strutturata (concetti nativi di classe proprietà ecc.)
 - Grande espressività

... Rappresentazione della conoscenza (KR)

- Logica del primo ordine (FOL) o suoi frammenti
 - Non decidibile se presa tutta
 - Nessuna struttura sui domini
- Logiche Descrittive (DL, DLs)
 - Spesso intrattabile
 - Difficile modellare relazioni m-n

Apprendimento da esempi

- Robusto su FOL (ad esempio Inductive Logic Programming)
- Non del tutto esplorato in DL
 - Kietz ed altri hanno provato a usare ILP per apprendere DL trasformando $DL \rightarrow FOL$
 - Kuesters ed altri usano approcci puri per apprendere in DL

KR per il Semantic Web

- DAML+ OIL è un istanza di una particolare DL chiamata SHIQ
- RDF è il linguaggio usato per esprimere asserzioni (\Leftrightarrow fatti in FOL)

Rimane da valutare l'efficacia dei metodi di apprendimento

Confronto preliminare

- Approccio ILP – oriented (DLP)
 - PAC learnability di alcune DL
 - Richiede trasformazione in FOL con aggiustamenti (perde vantaggio strutturazione)
 - Alcuni costrutti intraducibili (cfr. lavoro di Borgida)
- Approccio puro
 - Alcune operazioni intrattabili per alcune DL

Cosa abbiamo visto

- Semantic Web
 - Nuova generazione del Web
 - Integrabile con il precedente WWW
 - Ragionamento inserito nel WWW
- Architettura a Livelli
 - Sintattico
 - Metadati
 - Ontologie
 - Logico
 - Prova e fiducia

... Cosa abbiamo visto ...

- Tecnologie implementate
 - XML^{W3C} (interoperabilità sintattica)
 - RDF^{W3C} rappresentazione standard dei metadati
 - OWL^{W3C} et similia vocabolari condivisi e ontologie
 - RuleML, DAMLRules, OWLRules regole

... Cosa abbiamo visto ...

- Sfide
 - Non ri- scrivere il Web
 - Applicazioni che elaborano il significato delle informazioni
 - Informazioni estratte dal testo →
Ontologie e regole → Conoscenza

... Cosa abbiamo visto

- Metodi e Tecniche necessari
 - Natural Language Processing
 - Knowledge Management & Representation
 - First Order Logic
 - Description Logic
 - Reasoning
 - Theorem Proving
 - Reasoning
 - Apprendimento Automatico
 - Non Supervisionato
 - Supervisionato