

Linguaggi

*Corso M-Z - Laurea in Ingegneria Informatica
A.A. 2008-2009*

Alessandro Longheu

<http://www.diiit.unict.it/users/alongheu>

alessandro.longheu@diiit.unict.it

- *lezione 23* -

Semantic Web: Introduzione e linguaggi

1

A. Longheu – Linguaggi M-Z – Ing. Inf. 2008-2009

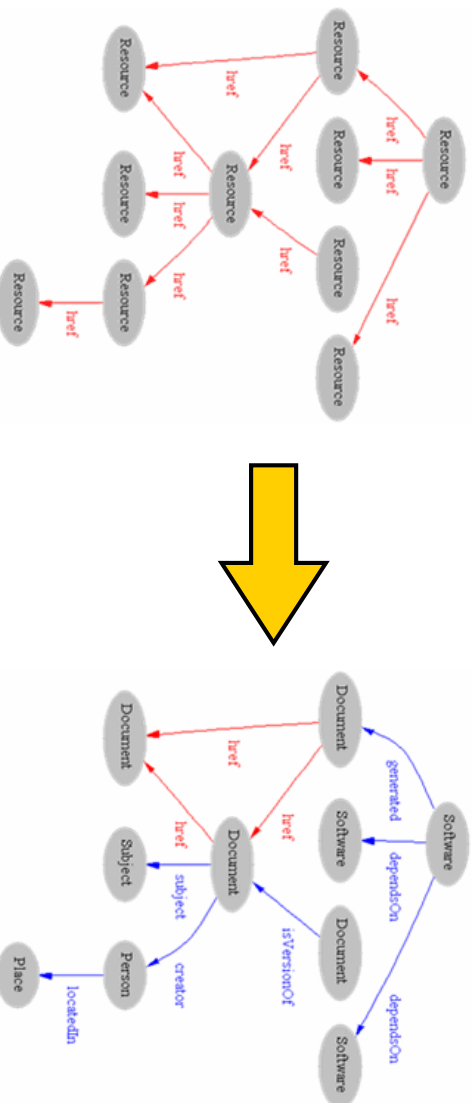
Web Semantico

- Con il termine Web Semantico si intende la **trasformazione** del World Wide Web in un ambiente dove i documenti pubblicati (pagine HTML, file, immagini, e così via) siano associati ad informazioni e dati (metadati) che ne specificano il contesto semantico in un formato adatto all'interrogazione, all'interpretazione e, più in generale, all'elaborazione automatica.
- Con l'interpretazione del contenuto dei documenti che il Web Semantico propugna, saranno possibili ricerche molto più evolute delle attuali, basate sulla presenza nel documento di parole chiave, ed altre operazioni specialistiche come la costruzione di reti di relazioni e connessioni tra documenti secondo logiche più elaborate del semplice link ipertestuale, permettendo un approccio simile a quello presente nei *sistemi esperti*

2

Web Semantico

- Le pagine web sono oggi collegate sintatticamente mediante indici che localizzano la URL della pagina. Uno dei principali limiti di tale impostazione risiede nell'assenza di significato dei collegamenti: i collegamenti dovrebbero anche descriverci il luogo in cui saremmo condotti (link semantico).



3

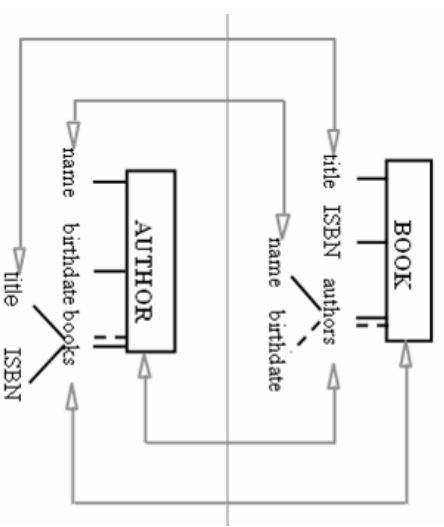
Web Semantico

- Il **WS non implica una qualche forma di intelligenza** perché Invece di richiedere ai computer di comprendere il linguaggio umano e la sua logica, **si richiede all'uomo di fare uno sforzo in più in fase di progettazione web.**
- Questo sforzo renderebbe il web **machine-understandable** e non semplicemente machine-readable, e risolverebbe uno dei problemi cronici dei motori di ricerca, quello del vocabolario, ad esempio casi di sinonimia e polisemia che rendono praticamente impossibile per i motori di ricerca restituire esclusivamente i risultati attesi, questo a causa della notevole ricchezza ed ambiguità) del linguaggio naturale, ad esempio la parola albero riguarda informatica, botanica, nautica? e ancora, un documento che parla di finanziamento del governo alle società calcistiche in pericolo di fallimento in che ambito ricade? Sport, politica, finanza?

4

Web Semantico

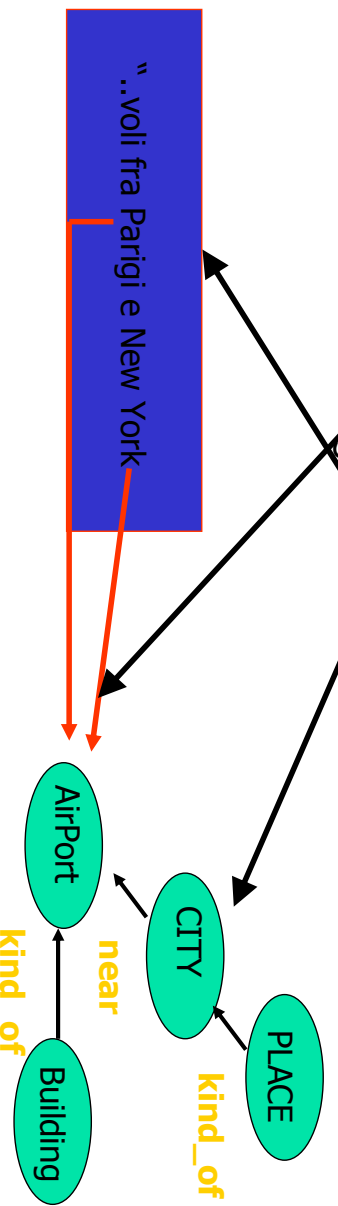
- Lo scenario futuro cerca di **riprodurre sul Web quello che già in parte esiste nel mondo dei database**: quando interroghiamo una base di dati, possiamo infatti fare ricerche piuttosto raffinate, ad esempio, chiedere “quali autori hanno scritto almeno due libri sull’IR”
- L’utente può formulare una richiesta che imponga precise relazioni (“almeno due libri sull’IR”), e tali relazioni sono stabilite fra concetti (“autore” e “libro”) non fra parole chiave (non si ricerca la stringa “autore” o “libro”). Questo è possibile perché esiste uno **schema del DB**, cioè un modello ed un insieme di regole che stabiliscono come debbano essere organizzati i dati




5

Web Semantico

- Nel web, invece, le informazioni sono in genere NON strutturate; è quindi necessario fornire tale struttura ai dati (le pagine web) tramite:
 - I metadati (HTML) o annotazioni (XML, RDF) per indicare i collegamenti semantici
 - Lo schema (o ontologia) del dominio per ragionare su tali collegamenti, estraendo le informazioni di interesse e/o trovando nuovi collegamenti semantici



6



Web Semantico

- Il linguaggio HTML fornisce una soluzione primitiva al problema della descrizione semantica dei contenuti utilizzando il tag META (che sta per meta-dato) nella sezione HEAD. I tag meta non vengono visualizzati a chi fruisce della pagina HTML come lettore bensì da taluni motori di ricerca per valutare l'appropriatezza di una pagina come risposta ad una certa domanda.
- In generale un tag meta ha la seguente forma:
- `<meta NAME="nome" CONTENT="valore">`
- Con questa indicazione l'autore di una pagina associa la stringa "valore" alla stringa "nome", dove "nome" indica una proprietà del documento e "valore" il valore di tale proprietà.

7



Web Semantico

- Esempi di utilizzo di metatag sono:
- `<head>`
- `<meta NAME="AUTHOR" CONTENT="C. Bianchi">`
- `<meta NAME="KEYWORDS" CONTENT="XML, programmazione web, query">`
- `<meta NAME="DESCRIPTION" CONTENT="introduzione all'XML">`
- `</head>`
- L'idea è quindi trasferire nella sezione HEAD le meta-informazioni relative ad un documento, privilegiando il contenuto di tale sezione come informazione utile durante una ricerca.

8

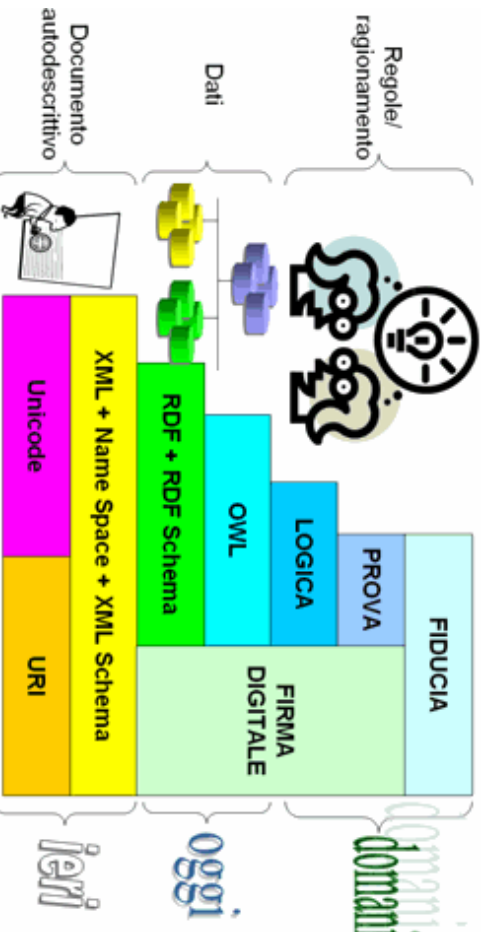
Web Semantico

- L'utilizzo del tag HTML <meta> presenta un primo limite, ossia non esistono convenzioni né sui nomi né sulla struttura del contenuto delle proprietà da definire, limitando fortemente l'utilizzo di sistemi automatici.
- Se un sistema non è in grado di identificare o distinguere nomi di proprietà non potrà, per esempio, distinguere il caso in cui le meta-informazioni specificano che il sito è stato scritto da "C. Bianchi" da quello in cui si dichiara che il sito riguarda la vita e le opere di "C. Bianchi"
- Il secondo limite dell'uso del tag <meta> è che non permette di esprimere legami semantici.

9

Web Semantico

- L'architettura del semantic web, necessaria allora per permettere ai dati di essere machine-understandable, ha diversi livelli:



10



Web Semantico

- URL vs. URN vs. URI
 - The difference between the three is subtle. An URL refers to a Web page, including the scheme, but without a name location. An URN may also include the location of a code fragment. An URI refers to a Web page including the location of the code fragment, if one exists, and the scheme.
 - URL <http://www.cnn.org/iis/review1.htm>
 - URN <www.cnn.org/iis/review1.htm#one>
 - URI <http://www.cnn.org/iis/review1.htm#one>
 - Because Web servers allow for default documents and do not require a scheme to retrieve a document, the subtle difference between an URL, URN and URI is hard to tell.


11



Web Semantico


- **XML, Name Space e XML Schema:**
 - XML (eXtensible Markup Language) è un meta-linguaggio di markup. In pratica fornisce un insieme di regole sintattiche per modellare la struttura di documenti e dati. Questo insieme di specifiche definiscono le modalità con cui crearsi un proprio linguaggio di markup. XML reca tra i suoi vantaggi fondamentali quello di garantire un'alta interoperabilità dei dati (e dunque di consentire l'interscambio di dati tra piattaforme ed applicativi diversi).
 - La struttura e la grammatica soggiacenti ad un documento XML possono essere stabilite attraverso un DTD (Document Type Definition) o (meglio) attraverso XML Schema, che fornisce un metodo per comporre vocabolari XML.
 - Un Namespace non è altro che un insieme di nomi di elementi e/o attributi identificati in modo univoco da un identificatore. La presenza di un identificatore univoco individua così un insieme di nomi distinguendoli da eventuali omonimie presenti in altri namespaces.

12



Web Semantico

- **RDF e RDF Schema:**
 - RDF (Resource Description Framework) fornisce un insieme di regole per definire informazioni descrittive sui dati, più precisamente sugli elementi costitutivi un documento web; queste asserzioni sono realizzate tramite triple che legano tra loro gli elementi in una relazione binaria. Le triple sono del tipo: Soggetto (la risorsa), Predicato (la proprietà) e Oggetto (il valore). Un modello RDF è rappresentabile da un grafo orientato sui cui nodi ci sono risorse o tipi primitivi e i cui archi rappresentano le proprietà.
 - RDF Schema fornisce, a sua volta, un metodo per combinare queste descrizioni in un singolo vocabolario. Il modo per sviluppare vocabolari specifici per un dato dominio di conoscenza è rappresentato dalle ontologie. 13



Web Semantico

- Nell'ambito del Web Semantico, il W3C ha sostenuto lo sviluppo di **OWL (Web Ontology Language)** quale linguaggio per la definizione di ontologie strutturate basate sul Web.
- OWL è un linguaggio di markup per rappresentare esplicitamente significato e semantica di termini con vocabolari e relazioni tra i termini. Tale rappresentazione dei termini e delle relative relazioni costituisce un'ontologia.
- L'obiettivo è permettere ad applicazioni software di elaborare il contenuto dei documenti scritti in OWL.



Web Semantico

- Logica, Prova e Fiducia:
 - **Logica:** Affinché il Web Semantico possa effettivamente aiutarci in una vasta gamma di situazioni, estraendo autonomamente informazioni utili dalla mole di documenti annotati semanticamente, è indispensabile costruire un potente linguaggio logico per realizzare le inferenze (ovvero procedimenti deduttivo mediante cui, a partire da una o più premesse, si ricava, per via logica, una conclusione).
 - **Prova:** Le conclusioni ottenute saranno validate a questo livello tramite motori di validazione costituiti da sequenze di formule derivate da assiomi.
 - **Trust:** Infine il sistema restituirà solo quelle informazioni che secondo il richiedente proverranno da utenti di indubbia attendibilità.

15



Web Semantico

- Volendo semplificare il discorso possiamo dire che alla base vi dev'essere una diversa e più attenta filosofia di progettazione delle risorse web - basate su XML -, le quali devono rispettare gli standard definiti e recare con se una descrizione delle proprie caratteristiche (tramite RDF e metadati).
- Ciascuna di queste risorse sarà identificabile in modo non ambiguo grazie all'uso degli URI (risolvendo così i problemi di ambiguità visti quando abbiamo parlato dei motori di ricerca).
- I metadati sono la base informativa su cui potranno operare gli agenti intelligenti per prendere le proprie decisioni.
- Gli agenti, a loro volta, potranno muoversi nello spazio-web sfruttando il sistema di rappresentazione della conoscenza disponibile (ontologie). Le decisioni degli agenti a questo punto saranno consentite grazie all'utilizzo di linguaggi di inferenza logica. Gli agenti, infine, nel prendere le proprie decisioni terranno conto del grado di fiducia attribuito alle risorse (ed ai loro autori identificati da sistemi di firma digitale) dagli utenti stessi.

16



Web Semantico

- La piena realizzazione dei principi del Web Semantico è probabilmente ancora lontana da una sua realizzazione e gli ostacoli maggiori al suo sviluppo si incontrano proprio al livello ontologico dell'architettura precedentemente vista.
- L'onerosità della mappatura delle risorse, la piena interoperabilità tra i diversi linguaggi utilizzati per la descrizione dei dati e le relazioni tra essi, i cambiamenti, anche culturali, profondi che si richiedono soprattutto in fase di progettazione dei documenti destinati al web richiedono uno sforzo supplementare e quell'adeguamento sociale e tecnologico che fin dagli inizi Berners Lee aveva indicato come chiave del cambiamento.

17



RDF

- **Resource Description Framework** (RDF) è una particolare *applicazione XML* (standardizzata dal W3C) deputata alla gestione delle **relazioni tra informazioni** ispirandosi ai principi della *logica dei predicati* (o logica predicativa del primo ordine) e ricorrendo agli strumenti tipici del Web (ad es. URI) e dell'XML (namespace).
- In estrema sintesi, secondo la logica dei predicati le informazioni sono esprimibili con **asserzioni** (statement) costituite da triple formate da **soggetto, predicato e valore** (noti anche come subject, verb e object, rispettivamente).
- RDF lavora quindi con tre tipi di elementi fondamentali:
 - **Risorse**: entità riferite attraverso URI
 - **Proprietà**: relazioni binarie tra risorse e/o valori atomici di tipo primitivo
 - **Affermazioni**: specificano il valore di una certa proprietà relativa a una risorsa

18