

RDF per la rappresentazione della conoscenza

Oreste Signore

Ufficio Italiano W3C¹, presso il C.N.R. - Istituto CNUCE
Area della Ricerca di Pisa San Cataldo - Via G. Moruzzi, 1 - 56124 Pisa
Email:oreste@w3.org

Abstract. Uno degli obiettivi a lungo termine del World Wide Web Consortium, che definisce regole e strategie di evoluzione del Web, è il Semantic Web. Una esigenza fondamentale per il raggiungimento di questo obiettivo è la rappresentazione della conoscenza in maniera comprensibile alla macchina. In questo lavoro verranno illustrati il ruolo dei metadati e le caratteristiche fondamentali di RDF.

Introduzione

Il World Wide Web Consortium (W3C, <http://www.w3.org>), è un consorzio che sviluppa tecnologie (specifiche, linee guida, software, e strumenti) per portare il Web al massimo del suo potenziale, definendo protocolli comuni che ne favoriscano l'evoluzione e assicurino l'interoperabilità. Gli obiettivi a lungo termine del W3C sono coerenti con le motivazioni iniziali che hanno portato alla nascita del web. Essi possono essere espressi sinteticamente come:

- *Universal Access*: Rendere il Web accessibile a tutti, promuovendo tecnologie che tengono conto delle notevoli differenze in termini di cultura, formazione, capacità, risorse materiali, e limitazioni fisiche degli utenti in tutti i continenti
- *Semantic Web*: Sviluppare un ambiente software che consenta ad ogni utente di fare il miglior uso possibile delle risorse disponibili sul Web
- *Web of Trust*: guidare lo sviluppo del Web tenendo in attenta considerazione gli aspetti innovativi che questa tecnologia solleva in campo legale, commerciale e sociale

Nel seguito, vedremo come la rappresentazione della conoscenza costituisca un elemento fondamentale per il raggiungimento di questi obiettivi, e descriveremo in particolare il ruolo dei metadati e la loro descrizione mediante RDF.

¹ Il W3C, guidato da Tim Berners-Lee, Direttore e creatore del World Wide Web, è stato costituito nell'ottobre 1994 con lo scopo di sviluppare al massimo il potenziale del World Wide Web. È un consorzio internazionale di imprese, neutrale rispetto ai venditori, ospitato congiuntamente da Massachusetts Institute of Technology Laboratory for Computer Science (MIT/LCS) negli Stati Uniti, Institut National de Recherche en Informatique et en Automatique (INRIA) in Europa, Keio University Shonan Fujisawa Campus in Giappone.

1. I metadati²

Nel navigare sul web, si seguono dei link, che portano a quella che formalmente viene detta *risorsa* (*resource*) identificata univocamente da un URI³. Nel linguaggio corrente una risorsa viene anche detta “*documento*”, per mettere in evidenza il fatto che sia leggibile da un essere umano, o “*oggetto*”, per mettere in evidenza che è leggibile da una macchina. Qualunque sia il termine utilizzato, la risorsa non è una entità a sé, ma è accompagnata da informazioni che la descrivono. Le informazioni sulla risorsa vengono generalmente dette *Metadati*.

Si può quindi dire che *i metadati sono informazioni, comprensibili dalla macchina, relative a una risorsa web o a qualche altra cosa*. Il punto chiave è costituito appunto dal fatto che i metadati sono comprensibili dalla macchina (*machine understandable*). Di conseguenza, i metadati costituiscono un tipo di informazione che può essere utilizzata dai *software agent*, per fare un uso appropriato delle risorse, rendendo più semplice e veloce il funzionamento del Web, aumentando la nostra fiducia in esso. A titolo di esempio, quando si reperisce un documento (o un oggetto) sul web, utilizzando il protocollo HTTP, è possibile che il server invii alcune informazioni sulla risorsa, quali la sua data di aggiornamento, la data massima di validità dell'informazione, il suo autore, etc. Quindi il Web, come insieme di risorse e di informazioni sulle risorse (cioè metadati) è già una realtà alla quale siamo abituati. Va tenuto presente che *i metadati sono dati*, e questo fatto ha alcune conseguenze.

La prima è che i metadati possono essere memorizzati come dati, in una risorsa, che può quindi contenere informazioni relative a se stessa o ad un' altra risorsa. Attualmente esistono tre modi per acquisire i metadati:

1. i metadati sono contenuti nella risorsa medesima, come per esempio nella sezione HEAD di un documento HTML;
2. al momento del trasferimento della risorsa, le informazioni vengono trasferite dal server al client (GET) o dal client al server (PUT o POST);
3. i metadati vengono estratti da un' altra risorsa;

quindi: *i metadati relativi ad un documento possono essere estratti dalla risorsa stessa, o da un' altra risorsa, o possono essere trasferiti con il documento*.

La seconda conseguenza del fatto che i metadati sono dati, è che *i metadati possono essere descritti da altri metadati*, e così via.

I metadati consistono in *asserzioni* sui dati, le quali vengono quindi rappresentate sotto forma di un *nome di asserzione* e un insieme di *parametri*. L' assioma è che *i metadati sono rappresentati da un insieme di asserzioni indipendenti*. Nel caso vi siano più asserzioni relative alla stessa risorsa, l' asserzione risultante è quella ottenuta mediante l' AND logico delle due asserzioni.

Le asserzioni relative alle risorse vengono spesso riferite come attributi della risorsa. Per esempio, se per una risorsa vale l' asserzione che essa ha una proprietà

² La descrizione ricalca in gran parte quella contenuta in [TBL1997]

³ **URI** (*Uniform Resource Identifier*). è il generico insieme di tutti i nomi/indirizzi che costituiscono le brevi sequenze di caratteri che fanno riferimento ad una risorsa.

URL (*Uniform Resource Locator*) è un termine informale, non più utilizzato nelle specifiche tecniche, associato con gli schemi URI più noti e diffusi (http, ftp, mailto, etc.).

specifica come l' autore, il parametro è il nome dell' autore. Analogamente, se il soggetto è un attributo della risorsa, allora il parametro è l' argomento trattato.

Gruppi di asserzioni relative alla stessa risorsa prendono spesso la forma di una lista di coppie (attributo-valore). Il termine asserzione enfatizza il fatto che la coppia (attributo-valore), quando viene trasferita con la risorsa, è uno statement fatto da una terza parte. Questo aspetto assume una particolare rilevanza nel contesto in cui il "Web of trust" diventa un elemento importante, in quanto ci permette di conoscere l' origine e l' affidabilità dei dati e dei metadati.

2. Il Resource Description Framework

Abbiamo già sottolineato come fosse difficile automatizzare il Web restando ancorati alla sua architettura originaria, in cui tutte le informazioni erano *machine-readable*, ma non *machine-understandable*, e come la soluzione al problema sembri venire dai *metadati*.

L' uso efficace dei metadati, tuttavia, richiede che vengano stabilite delle convenzioni per la *semantica*, la *sintassi* e la *struttura*. Le singole comunità interessate alla descrizione delle loro risorse specifiche definiscono la semantica dei metadati pertinenti alle loro esigenze. La sintassi, cioè l' organizzazione sistematica dei data element per l' elaborazione automatica, facilita lo scambio e l' utilizzo dei metadati tra applicazioni diverse. La struttura può essere vista come un vincolo formale sulla sintassi, per una rappresentazione consistente della semantica.

RDF (Resource Description Framework) è lo strumento base per la codifica, lo scambio e il riutilizzo di metadati strutturati, e consente l' interoperabilità tra applicazioni che si scambiano sul Web informazioni machine-understandable. I settori nei quali RDF può essere utilizzato e portare vantaggi sono i più disparati, basti citare, a titolo di esempio:

- descrizione del contenuto di un sito Web, o di una pagina, o di una biblioteca digitale;
- implementazione di *intelligent software agent*, per lo scambio di conoscenza e un utilizzo migliore delle risorse Web;
- *classificazione* del contenuto, per applicare criteri di selezione;
- descrizione di un *insieme di pagine*, che rappresentano un singolo documento logico;
- stabilire i criteri di *proprietà intellettuale* delle singole pagine;
- esprimere criteri di *privacy preference* degli utenti e le *privacy policies* di un sito Web;
- con il meccanismo della *digital signature*, contribuire alla creazione del Web of Trust, per le applicazioni nel commercio elettronico, la cooperazione, etc..

Il **Resource Description Framework** (RDF), quindi, non descrive la semantica, ma fornisce una base comune per poterla esprimere, permettendo di definire la semantica dei tag XML. RDF è costituito da due componenti:

- **RDF Model and Syntax**: definisce il *data model* RDF e la sua codifica XML;
- **RDF Schema**: permette di definire specifici *vocabolari* per i metadati.

3. RDF Data Model

RDF fornisce un modello per descrivere le risorse. Come abbiamo visto precedentemente, le risorse hanno delle proprietà (o anche attributi o caratteristiche). RDF definisce una risorsa come un qualsiasi oggetto che sia identificabile univocamente mediante un Uniform Resource Identifier (URI).

Il data model RDF è molto semplice, ed è basato su tre tipi di oggetti:

Resources Qualunque cosa descritta da una espressione RDF viene detta risorsa (*resource*). Una risorsa può essere una pagina Web, o una sua parte, o un elemento XML all'interno del documento sorgente. Una risorsa può anche essere un'intera collezione di pagine web, o anche un oggetto non direttamente accessibile via Web (per es. un libro, un dipinto, etc.). Le risorse sono sempre individuate da un URI, eventualmente con un anchor id.

Properties Una *property* (proprietà) è un aspetto specifico, una caratteristica, un attributo, o una relazione utilizzata per descrivere una risorsa. Ogni proprietà ha un significato specifico, definisce i valori ammissibili, i tipi di risorse che può descrivere, e le sue relazioni con altre proprietà. Le proprietà associate alle risorse sono identificate da un *nome*, e assumono dei *valori*.

Statements Una risorsa, con una proprietà distinta da un nome, e un valore della proprietà per la specifica risorsa, costituisce un *RDF statement*. Uno statement è quindi una tupla composta da un *soggetto* (risorsa), un *predicato* (proprietà) e un *oggetto* (valore). L'oggetto di uno statement (cioè il property value) può essere un'espressione (sequenza di caratteri o qualche altro tipo primitivo definito da XML) oppure un'altra risorsa.

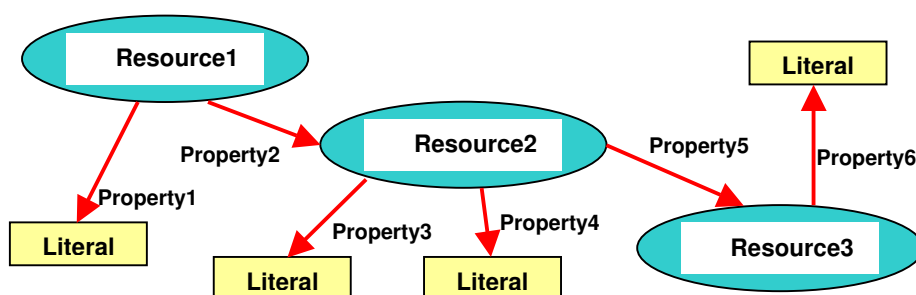


Fig. 1. Il modello generico di una descrizione RDF

Graficamente, le relazioni tra Resource, Property e Value vengono rappresentate mediante *grafi etichettati orientati*, in cui le risorse vengono identificate come nodi (graficamente delle ellissi), le proprietà come archi orientati etichettati, e i valori corrispondenti a sequenze di caratteri come rettangoli. Una rappresentazione grafica di una generica descrizione RDF è quella della figura 1.

Un insieme di proprietà che fanno riferimento alla stessa risorsa viene detto descrizione (*description*).

3.1 Qualche esempio

L' utilizzo di RDF può essere chiarito con qualche semplice esempio.

Consideriamo queste due espressioni:

1. "Oreste Signore è l' autore del DocumentoX"
2. "L' autore del DocumentoX è Oreste Signore"

Ovviamente, le due espressioni sono del tutto equivalenti per un essere umano, in quanto veicolano la stessa informazione, mentre verrebbero viste come due espressioni diverse da una macchina. RDF, mediante il suo semplice modello basato su resource, property e value, intende fornire un metodo non ambiguo per esprimere la semantica con una codifica comprensibile dalla macchina.

Dato che RDF fornisce un meccanismo per associare le proprietà alle risorse, per poter dare un valore alle proprietà occorre che sia dichiarata la risorsa. Quindi, per prima cosa va dichiarata una risorsa che rappresenti il DocumentoX. Abbiamo quindi la tripla:

<i>Resource</i>	<code>http://www.w3c.it/Oreste/DocX</code>
<i>Property</i>	author
<i>Value</i>	Oreste Signore

Lo statement dell' esempio verrebbe quindi rappresentato come:

La risorsa `http://www.w3c.it/Oreste/DocX` has Author Oreste Signore

E verrebbe rappresentata graficamente come in Figura 2.



Fig. 2. La rappresentazione grafica di uno statement RDF

La direzione dell' arco è importante: essa parte dal soggetto dello statement e punta all' oggetto.

Se invece vogliamo dire qualcosa in più riguardo all' oggetto, per esempio vogliamo dire:

Oreste Signore, la cui Email è `oreste@w3.org`, e lavora presso il C.N.R., è l' autore del DocumentoX

desideriamo esprimere delle informazioni *riguardo a (about)* Oreste Signore, e quindi trasformare il valore della proprietà autore in una entità strutturata. In RDF, un' entità di questo tipo viene rappresentata come un' altra risorsa. Poiché nello statement non

viene attribuito un nome a questa entità strutturata, il diagramma risultante sarebbe quello di Figura 3.

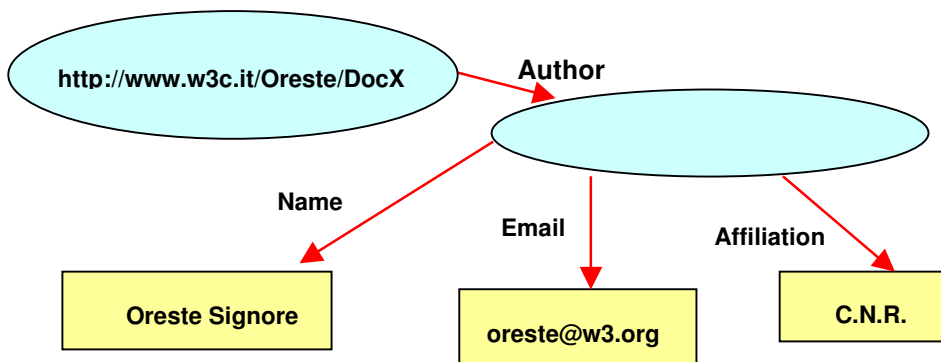


Fig. 3. Una proprietà con un valore strutturato

Si noti che questo diagramma potrebbe essere letto come:

http://www.w3c.it/Oreste/DocX has author qualcuno e questo qualcuno has Name Oreste Signore, Email oreste@w3.org, e Affiliation C.N.R.

Alla risorsa anonima “qualcuno” potrebbe essere assegnato un identificatore univoco, per esempio il CodiceFiscale, corrispondente a un URI del tipo:

<http://www.finanze.it/CF/SGNRST99A99X111Y>.

Il diagramma risultante sarebbe quello di Figura 4.

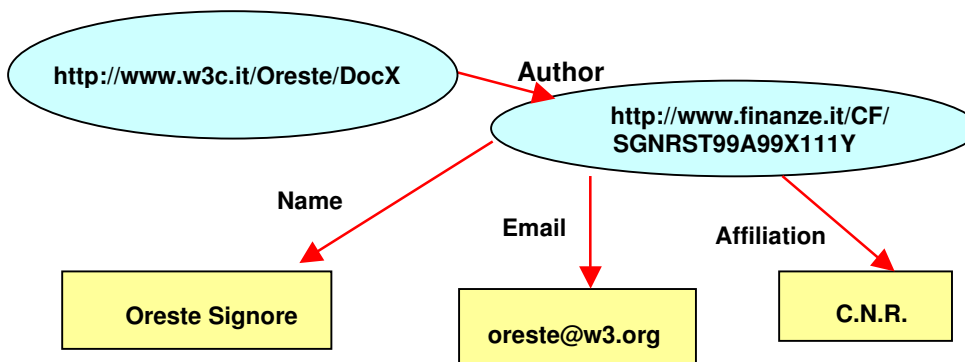


Fig. 4. Un valore strutturato con identificatore

Questo diagramma potrebbe essere letto come la concatenazione di due frasi:

La persona identificata dal Codice Fiscale SGNRST99A99X111Y has Name Oreste Signore, Email oreste@w3.org, e Affiliation C.N.R..

La risorsa http://www.w3c.it/Oreste/DocX has author questa persona

In questo esempio è stata creata una risorsa, identificabile univocamente, per l' autore, ma non per il nome, la Email, l' affiliazione. Il modello RDF consente la creazione di risorse a più livelli. Per esempio, sarebbe stato possibile creare una risorsa per l' Affiliazione, con proprietà come: tipoDiEnte, partita IVA, sedeSociale, etc. Quali siano i limiti pratici e logici per il numero di livelli, dipende essenzialmente dalle caratteristiche e dalle tradizioni delle singole comunità che definiscono la descrizione delle risorse.

3.2 I Container

Talvolta, è necessario far riferimento a più di una risorsa, per esempio per descrivere il fatto che un libro è stato scritto da più autori, oppure che un documento è composto da una serie di componenti, oppure che una funzione può essere svolta da una delle persone elencate. RDF definisce tre tipi di contenitori (*container*):

Bag È una lista non ordinata di risorse o costanti. Viene utilizzato per dichiarare che una proprietà ha valori multipli, senza alcun significato particolare attribuito al loro ordine (per esempio, i componenti di una commissione). Sono ammessi valori duplicati.

Sequence È una lista ordinata di risorse o costanti. Viene utilizzato per dichiarare che una proprietà ha valori multipli, e che il loro ordine è significativo (per esempio, gli autori di un libro, un insieme di nomi di cui si voglia preservare l' ordine alfabetico). Sono ammessi valori duplicati.

Alternative È una lista di risorse o costanti che rappresentano una alternativa per il valore (singolo) di una proprietà. Può essere utilizzato, per esempio, per fornire titoli alternativi in varie lingue.

È possibile definire proprietà sia dell' intero container che dei singoli elementi.

3.3 Container e proprietà multiple

Una risorsa può essere soggetto in più statement, sempre con lo stesso predicato (per esempio, *Calvino è autore di "Se una notte d' inverno un viaggiatore", "Le fiabe italiane", "Il barone rampante"*). Si noti che è semanticamente diverso il caso in cui si ha un singolo statement il cui oggetto è un container contenente vari esemplari. Per esempio, lo statement: *"La commissione composta da X, Y e Z ha adottato una decisione"*, non implica che ogni membro della commissione abbia espresso lo stesso parere, come invece sarebbe nel caso in cui si usasse uno statement multiplo.

3.4 Statements about statements

Gli statement riguardano le risorse. In alcuni casi, è utile poter certificare la credibilità di un particolare statement, quindi formulare degli statement relativi ad altri statement. Per esempio:

La risorsa <http://www.w3c.it/Oreste/DocX> has Author Oreste Signore

Viene vista da RDF come un fatto. Invece, lo statement:

*Carlo Sorge dice che la risorsa <http://www.w3c.it/Oreste/DocX> has
Author Oreste Signore*

Non afferma un fatto relativo alla risorsa <http://www.w3c.it/Oreste/DocX>, ma un fatto relativo all'affermazione di Carlo Sorge. Per esprimere questo fatto a RDF, dobbiamo modellare lo statement come una risorsa con quattro proprietà. Nell'ambito della comunità che si interessa di Knowledge Representation, questo processo prende il nome di *reifificazione (reification)*, e il modello di statement prende il nome di *reified statement*.

Per modellare gli statement, RDF definisce queste quattro proprietà sono:

- subject* identifica la risorsa che viene descritta dallo statement modellato, quindi il soggetto è la risorsa relativamente alla quale era stato formulato lo statement originale (<http://www.w3c.it/Oreste/DocX>, nell'esempio).
- predicate* identifica la proprietà originale nello statement modellato. Il valore del predicato è una risorsa che rappresenta la specifica proprietà nello statement originale (nel nostro esempio, Author).
- object* identifica il valore della proprietà nello statement modellato. Il valore di object è l'object nello statement originale (nel nostro esempio: "Oreste Signore")
- type* descrive il tipo della nuova risorsa

Una nuova risorsa con queste quattro proprietà rappresenta lo statement originale, e può essere utilizzata come object di altri statement e avere ulteriori statement che lo riguardano.

3.5 I Namespace

RDF consente alle singole comunità di definire la semantica. Tuttavia, non è possibile affidare la semantica semplicemente al nome, che potrebbe avere significati più o meno ampi a seconda degli interessi specifici delle singole comunità. RDF identifica univocamente le proprietà mediante il meccanismo dei namespace. ([XMLns]). I namespace XML forniscono un metodo per identificare in maniera non ambigua la semantica e le convenzioni che regolano l'utilizzo delle proprietà identificando l'authority che gestisce il vocabolario.

Uno degli esempi più noti è la Dublin Core Initiative ([DC], [DCES], [DCRDF]) che definisce, per esempio, il campo "*Subject and Keywords*" nel seguente modo:

```
Name:      Subject and Keywords
Identifier: Subject
Definition: The topic of the content of the resource.
Comment: Typically, a Subject will be expressed as keywords,
         key phrases or classification codes that describe a
         topic of the resource.
```


Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.

Si può quindi utilizzare un namespace XML per identificare in maniera non ambigua lo schema per il vocabolario Dublin Core puntando alla risorsa Dublin Core che ne definisce la semantica.

Quindi, la descrizione di un sito Web mediante le proprietà definite nel vocabolario Dublin Core e quelle di una personale estensione potrebbe essere:

```
<rdf:RDF
  xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
  xmlns:dc="http://purl.org/metadata/dublin_core#"
  xmlns:mydc="http://www.w3c.it/metadata/DCaddendum#">
  <rdf:Description about="http://www.dlib.org">
    <dc:Title>
      D-Lib Program - Research in Digital Libraries
    </dc:Title>
    <dc:Description>
      The D-Lib program supports the community of people
      with research interests in digital libraries and
      electronic publishing
    </dc:Description>
    <dc:Publisher>
      Corporation For National Research Initiatives
    </dc:Publisher>
    <dc>Date>1995-01-07</dc>Date>
    <dc:Subject>
      <rdf:Bag>
        <rdf:li>Research; statistical methods</rdf:li>
        <rdf:li>Education, research, related topics</rdf:li>
        <rdf:li>Library use Studies</rdf:li>
      </rdf:Bag>
    </dc:Subject>
    <dc>Type>World Wide Web Home Page</dc>Type>
    <dc:Format>text/html</dc:Format>
    <dc:Language>en</dc:Language>
    <mydc:Rating>
      Well known and often referenced site
    </mydc:Rating>
    <mydc:Originality>High</mydc:Originality>
  </rdf:Description>
</rdf:RDF>
```

Si noti in questo esempio, che costituisce una variante di uno di quelli presentati in [RDFMSS], la presenza di tre *namespace*, referenziati dai prefissi **rdf**, **dc** e **mydc** che permettono di utilizzare le proprietà definite nei tre namespace. In particolare, il namespace mydc permette di ampliare il numero di proprietà definite dal namespace che referencia Dublin Core.

4. RDF Schema

Il data model RDF permette di definire un modello semplice per descrivere le relazioni tra le risorse, in termini di proprietà identificate da un nome e relativi valori. Tuttavia, RDF data model non fornisce nessun meccanismo per dichiarare queste proprietà, né per definire le relazioni tra queste proprietà ed altre risorse.

RDF Schema permette di definire dei vocabolari, quindi l'insieme delle proprietà semantiche individuata da una particolare comunità.

RDFSchema permette definire significato, caratteristiche e relazioni di un insieme di proprietà, compresi eventuali vincoli sul dominio e sui valori delle singole proprietà. Inoltre, implementando il concetto (transitivo) di classe e sottoclasse, consente di definire gerarchie di classi, con il conseguente vantaggio che agenti software intelligenti possono utilizzare queste relazioni per svolgere i loro compiti.

5. Uno sguardo al futuro del Web

La sfida del futuro è il *Semantic Web*. Nella visione di Berners-Lee, il Web ha una architettura a livelli (Figura 5), che verrà sviluppata completamente nel giro di vari anni.

Per chiarezza di terminologia, va ricordato che la filosofia di base del Web è quella di uno spazio informativo universale, navigabile, con un mapping da *URI (Uniform Resource Identifier)* alle risorse.

Nel contesto del Semantic Web, il termine semantico assume la valenza di "*elaborabile dalla macchina*" e non intende fare riferimento alla semantica del linguaggio naturale e alle tecniche di intelligenza artificiale. Il Semantic Web è, come l'XML, un ambiente dichiarativo, in cui si specifica il significato dei dati, e non il modo in cui si intende utilizzarli. La semantica dei dati consiste nelle informazioni utili perché la macchina possa utilizzarli nel modo corretto, eventualmente convertendoli. In questo senso, si potrebbe parlare del superamento di un "*semantic test*" quando una macchina, ricevendo dei dati, li utilizza nella maniera corretta.

Si riproduce quindi la situazione di una decina di anni fa, quando vi furono le prime dimostrazioni del Web: l'aspetto essenziale è il fattore di scala. Su piccola scala, un collegamento tra due informazioni semanticamente equivalenti, ma rappresentate in maniera diversa su due database, non rappresenta qualcosa di significativo, ma ampliare queste possibilità alla comunità mondiale, senza imporre vincoli stringenti, assume tutta un'altra rilevanza.

Esaminando più in dettaglio gli elementi del Semantic Web, appare evidente, oltre al ruolo di base giocato da XML (con *Name Space* e *xmlschema*), la centralità di RDF e RDF Schema, che costituiscono il linguaggio per descrivere le risorse e i loro tipi.

Al di sopra di questo strato si pone il *livello ontologico*. Una ontologia permette di descrivere le relazioni tra i tipi di elementi (per es. "questa è una proprietà transitiva") senza però fornire informazioni su come utilizzare queste relazioni dal punto di vista computazionale.

La firma digitale (*digital signature*) è di significativa importanza in diversi strati nel modello astratto del Semantic Web. La crittografia a chiave pubblica è una tecnica

nota da qualche anno, ma non si è ancora diffusa su larga scala come ci si poteva attendere. Nella visione di Berners-Lee, un elemento che potrebbe aver giocato contro la diffusione di questa tecnica è che essa è a "grana grossa", imponendo una scelta binaria tra fiducia o non fiducia (trusted/not trusted), mentre sarebbe necessaria una infrastruttura in cui le parti possano essere riconosciute e accettate come credibili in specifici domini. Con una granularità più fine come questa, la firma digitale potrebbe essere utilizzata per stabilire la provenienza delle ontologie e delle deduzioni, oltre che dei dati.

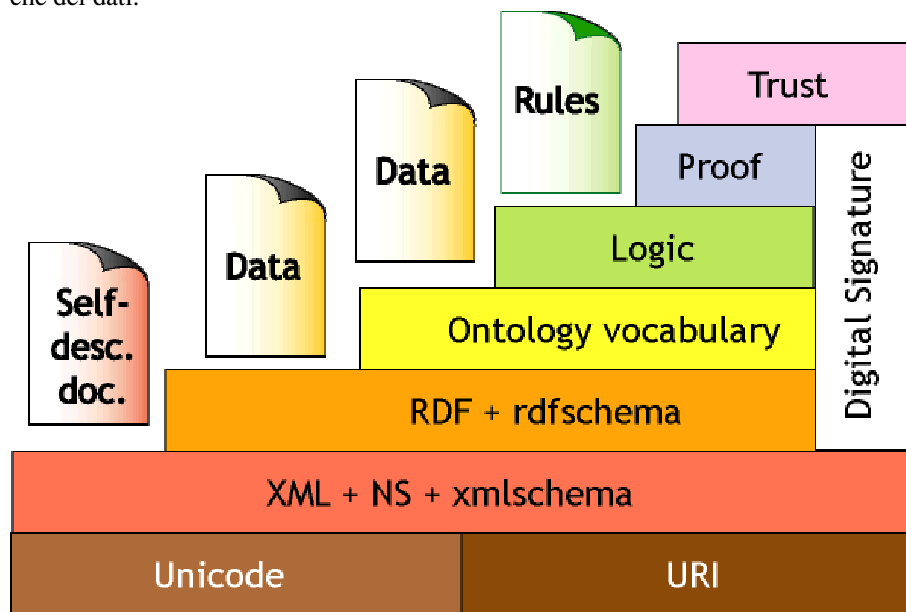


Fig. 5. L'architettura del Semantic Web
<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>

Il *livello logico* è il livello immediatamente superiore al livello ontologico. A questo livello le asserzioni esistenti sul Web possono essere utilizzate per derivare nuova conoscenza. Tuttavia, i sistemi deduttivi non sono normalmente interoperabili, per cui, secondo Berners-Lee, invece di progettare un unico sistema onnicomprensivo per supportare il ragionamento (*reasoning system*), si potrebbe pensare di definire un linguaggio universale per rappresentare le dimostrazioni. I sistemi potrebbero quindi autenticare con la firma digitale queste dimostrazioni ed esportarle ad altri sistemi che le potrebbero incorporare nel Semantic Web.

Anche se i livelli più alti dell'architettura potranno richiedere diversi anni prima di poter raggiungere uno stadio in cui siano effettivamente implementati e affidabili, già oggi esiste una quantità significativa di lavoro nel settore delle ontologie. Le soluzioni pratiche includono l'uso di XSLT per derivare RDF da documenti XML, la comparsa di database e motori di ricerca RDF di tipo generalizzato, interfacce grafiche generalizzate e specifiche per RDF.

L'intera comunità scientifica sta investendo molte energie nel settore della Semantic Web. Molti riferimenti utili si trovano in [SemWeb].

Conclusioni

Le linee di sviluppo del Web sono coordinate dal W3C, che le definisce con il supporto e l' accordo dei suoi membri. La tumultuosa crescita delle applicazioni Web e delle tecnologie relative impone ritmi quasi impensabili, come provato dalla quantità di documentazione tecnica che si sussegue con cadenza incredibile. In un clima di apparente confusione, esistono tuttavia alcune linee di sviluppo ben definite.

L' utilizzo di XML si va diffondendo con una rapidità e una pervasività che testimoniano come molti abbiano compreso che si tratta di una tecnologia, o meglio di una famiglia di tecnologie, che sono centrali nello sviluppo del Web.

In questo quadro si inserisce l' ambizioso progetto del Semantic Web, che, considerato da un certo punto di vista, tende a rendere reale il grande sogno di Ted Nelson ([NelsonTH], [LitMach1993]). Va sottolineato che un reale sviluppo di queste idee si potrà avere solo se si condividerà la filosofia ispiratrice del Web, mondo libero, che rifiuta o comunque riduce al minimo i vincoli della centralizzazione, delle soluzioni proprietarie, della costituzione e prevalenza di posizioni di potere.

In quest' ottica di sviluppo, uno dei problemi fondamentali è l' accesso all' informazione distribuita, che ora il Web propone a un livello di scala mai visto prima. I metadati portano un contributo fondamentale al miglioramento dell' accesso all' informazione, e RDF è lo standard proposto dal W3C per definire le architetture necessarie a supportare i metadati. RDF è una applicazione XML che permette di esprimere la semantica per la codifica, lo scambio e l' elaborazione automatica dei metadati. Inoltre, RDF consente di rendere disponibili, in una forma comprensibile sia ad un lettore umano che a una macchina, i vocabolari progettati per lo scambio, l' uso e l' estensione della semantica dei metadati tra comunità specialistiche diverse.

Riprendendo [Miller1998], possiamo descrivere l' impatto dell' adozione su vasta scala di RDF riportando quanto sostenuto da Ora Lassila in [IRDF]:

Once the web has been sufficiently "populated" with rich metadata, what can we expect? First, searching on the web will become easier as search engines have more information available, and thus searching can be more focused. Doors will also be opened for automated software agents to roam the web, looking for information for us or transacting business on our behalf. The web of today, the vast unstructured mass of information, may in the future be transformed into something more manageable - and thus something far more useful.

Bibliografia

- [Bray2001] Tim Bray: *What is RDF?*, URL: <http://www.xml.com/pub/a/2001/01/24/rdf.html>
- [DC] The Dublin Core Home Page, URL: <http://dublincore.org/>

- [DCES] Description of the Dublin Core Elements, URL:
<http://dublincore.org/documents/dces/>
- [DCRDF] OCLC News Release, "Dublin Core and Web MetaData Standards Converge in Helsinki", Nov. 7, 1997,
URL:<http://www.oclc.org/oclc/press/971107a.htm>
- [Dornfest2001] Rael Dornfest and Dan Brickley, *The Power of Metadata*, URL:
<http://www.openp2p.com/lpt/a/lp2p/2001/01/18/metadata.html>
- [Engelbart1990] Douglas C. Engelbart: *Knowledge-Domain Interoperability and an Open Hyperdocument System*, June 1990 (AUGMENT,132082,),
<http://www.bootstrap.org/augment-132082.htm>
- [IRDF] Introduction to RDF Metadata, W3C NOTE 1997-11-13, Ora Lassila,
URL:<http://www.w3.org/TR/NOTE-rdf-simple-intro>
- [Lassila1998] Ora Lassila: *Web Metadata: a Matter of Semantics*, IEEE Internet Computing, July/August 1998 (Vol. 2, No. 4), pp. 30-37, URL:
<http://computer.org/internet/ic1998/w4030abs.htm>
- [LitMach1993] Nelson Theodor Holm: *Literary Machines 93.1*,
<http://www.sfc.keio.ac.jp/~ted/TN/PUBS/LM/LMpage.html>
- [Miller1998] Miller E.: An Introduction to the Resource Description Framework, D-Lib Magazine, May 1998,
<http://www.dlib.org/dlib/may98/miller/05miller.html>
- [NelsonTH] <http://www.sfc.keio.ac.jp/~ted/>
- [RDFMSS] O.Lassila, R.Swick: *Resource Description Framework (RDF) Model and Syntax Specification*, W3C Recommendation 22 February 1999,
<http://www.w3.org/TR/REC-rdf-syntax>
- [RDFSS] *Resource Description Framework (RDF) Schema Specification*, W3C Recommendation 03 March 1999, <http://www.w3.org/TR/1999/PR-rdf-schema-19990303>
- [SemWeb] <http://www.semanticweb.org/>
- [TBL1997] Tim Berners-Lee: *Metadata architecture*, (1997),
<http://www.w3.org/DesignIssues/Metadata.html>
- [TBL1998] Tim Berners-Lee: *Semantic Web Road Map*, (1998),
<http://www.w3.org/DesignIssues/Semantic.html>
- [TBL1999] Tim Berners-Lee: *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*, HarperSanFrancisco (1999), ISBN 0-06-251587-X
- [TBL2001] Berners-Lee T., Hendler J., Lassila O.: *The Semantic Web*, Scientific American, May 2001,
<http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>
- [URI] T. Berners-Lee, R. Fielding, U.C. Irvine, L. Masinter: *Uniform Resource Identifiers (URI): Generic Syntax*; Internet Draft Standard August, 1998; [RFC2396](http://www.ietf.org/rfc/rfc2396.txt)
- [W3C] *World Wide Web Consortium Home Page*, <http://www.w3.org>
- [W3CML] World Wide Web Consortium (W3C) Members -
<http://www.w3.org/Consortium/Member/List>

- [W3CPD] World Wide Web Consortium Process Document - 19 July 2001
<http://www.w3.org/Consortium/Process-20010719/>
- [XML] *Extensible Markup Language (XML)* , <http://www.w3.org/XML/>
- [XML1.0] *Extensible Markup Language (XML) 1.0 (Second Edition) W3C Recommendation 6 October 2000*, <http://www.w3.org/TR/2000/REC-xml-20001006>
- [XMLns] *Namespaces in XML* - World Wide Web Consortium 14-January-1999
<http://www.w3.org/TR/REC-xml-names/>