

Una panoramica delle tecnologie W3C

Oreste Signore

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

Abstract. Nell' evoluzione del Web verso il Semantic Web, un ruolo primario viene giocato dalla famiglia di tecnologie XML. Già oggi è possibile sviluppare applicazioni che, utilizzando in modo coerente le tecnologie XML, possono conseguire livelli di significativi di accessibilità, portabilità e flessibilità con costi ridotti e salvaguardia degli investimenti. Inoltre, la coerenza con i principi che sono alla base del Web e con gli obiettivi a lungo termine del World Wide Web Consortium (W3C) garantisce una evoluzione delle applicazioni verso l' ambizioso obiettivo del Semantic Web. In questo lavoro vengono presentate sommariamente alcune specifiche W3C. Infine, vengono discusse le prospettive future, in particolar modo per quanto riguarda il Semantic Web.

Introduzione

Il World Wide Web Consortium (W3C, <http://www.w3.org>), è un consorzio che sviluppa tecnologie (specifiche, linee guida, software, e strumenti), definendo protocolli comuni che ne favoriscano l' *evoluzione* e assicurino l' *interoperabilità*. Gli *obiettivi a lungo termine* del W3C 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

La famiglia di tecnologie W3C basate su XML è estremamente vasta e in rapida evoluzione. Sul sito ([W3C]) è disponibile ampia documentazione, una sintesi molto stringata è in [Signore2001]. Nel seguito, verranno descritte brevemente alcune tecnologie W3C, illustrando l' aspetto *strutturale* (XML e XMLSchema) e *semantico*

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.

(metadati, RDF, Namespace), e le problematiche della *trasformazione* e fruizione dell' informazione (XSL e XSLT). Successivamente, verranno riportate alcune considerazioni sull' accessibilità e si farà un riferimento alle prospettive future (Semantic Web). In Appendice vengono riportate alcune informazioni relative al Consorzio e alle modalità di partecipazione alle sue attività.

1 - La strutturazione dell' informazione

1.1 - XML

Extensible Markup Language (XML) è nato per far fronte alle limitazioni di HTML nella realizzazione delle nuove applicazioni Web, in cui i dati costituiscono un elemento essenziale (*data-centric Web applications*). XML è stato quindi il primo passo per assegnare una semantica ai tag e supportare le transazioni sul Web, permettendo lo scambio di informazioni tra database diversi. Ulteriori e significativi vantaggi sono costituiti dalla possibilità di avere viste diverse degli stessi dati, e la possibilità di personalizzare le informazioni mediante opportuni agenti. L' adozione di XML agevola la gestione di collezioni di documenti, e costituisce un supporto fondamentale per la pubblicazione di informazioni a livello internazionale, con il non piccolo vantaggio di essere indipendente dalla piattaforma e dal linguaggio.

```
01 <?xml version="1.0"?>
02 <!DOCTYPE ordine [
03 <!ELEMENT ordine ( cliente, prodotto+ )>
04 <!ATTLIST ordine id ID #REQUIRED>
05 <!ELEMENT cliente EMPTY>
06 <!ATTLIST cliente db CDATA #REQUIRED>
07 <!ELEMENT prodotto ( importo )>
08 <!ATTLIST prodotto db CDATA #REQUIRED>
09 <!ELEMENT importo ( #PCDATA )>
10 ]>
11 <ordine id="ord001">
12   <cliente db="codcli123"/>
13   <prodotto db="prod345">
14     <importo>23.45</importo>
15   </prodotto>
16 </ordine>
```

Figura 1 - Un documento XML (con la sua DTD in grassetto)

Le caratteristiche di XML possono essere illustrate con un esempio semplice, relativo alla gestione degli Ordini. (Figura 1). La sintassi XML usa tag di inizio e fine, come per esempio `<importo>` e `</importo>`, per marcare i campi informativi. Un campo informativo racchiuso tra due marcatori viene detto elemento (*element*) e può essere ulteriormente arricchito dalla presenza di coppie nome/valore (nell' esempio, `id="ord001"`) dette attributi (*attribute*). Come si può vedere, si tratta di una sintassi semplice, la cui elaborazione automatica è poco complessa, senza codifiche particolarmente criptiche, per cui resta comprensibile alla lettura diretta. I tag devono essere inseriti correttamente uno dentro l' altro, deve esistere una corrispondenza tra il tag di apertura e quello di chiusura, sono previsti elementi a campo informativo nullo e gli attributi dei tag devono essere racchiusi tra doppi apici.

La presenza di una struttura formale del documento, espressa nella **DTD** (*Document Type Definition*), non ha un impatto diretto sul modello strutturale implicito: nell' esempio di Figura 1, in cui la DTD è inclusa nel documento (ma potrebbe anche essere referenziata come risorsa esterna) la riga 6 specifica che l' attributo *db* è obbligatorio. Un documento XML si dice "*well formed*" quando rispetta le regole di scrittura; viene detto "*validato*" quando è coerente con la struttura definita nella DTD.

XML, mezzo espandibile e flessibile per modellare il Web, costituisce attualmente la tecnologia chiave di W3C. XML ricopre un ruolo centrale nell' architettura del Web, e ogni nuovo linguaggio utilizzato per definire un nuovo standard deve essere descritto in XML. W3C considera XML come una famiglia di tecnologie, e non intende centralizzarne il controllo, preferendo lasciare agli utenti, coerentemente con la filosofia del Web, il compito di sviluppare applicazioni particolari.

1.2 - XML Schema Definition

La DTD presenta alcune limitazioni, riconducibili essenzialmente al fatto che viene espressa con una sintassi sua propria, e quindi richiede editor, parser e processor ad hoc. Inoltre, è difficile estenderla, non contempla datatype e deve supportare tutti gli elementi e attributi descritti dai namespace² inclusi.

Gli schema hanno le stesse funzionalità delle DTD, ma offrono alcuni significativi vantaggi: sono espressi con la sintassi XML e includono datatype, inheritance, regole di combinazione degli schema. **XMLSchema** fornisce anche un miglior supporto dei namespace e offre la possibilità di agganciare documentazione e informazioni semantiche. XMLSchema permette di rappresentare vincoli sui possibili valori, tipi complessi e gerarchie di tipi. In definitiva, gli XMLSchema sono strumenti molto più potenti delle DTD, e sul sito W3C sono disponibili parser, validatori, e altri strumenti utili. Utilizzare questa specifica nella realizzazione di nuove applicazioni costituisce un indubbio investimento per il futuro.

2 - La semantica

2.1 - I metadati³

Nel navigare sul web, si seguono dei link, che portano a quella che formalmente viene detta *risorsa* (*resource*) identificata univocamente da un URI⁴. 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

² Un **XML namespace** è un insieme di nomi, caratterizzati da un URI di riferimento, utilizzati come element type e attribute name. Il concetto sarà chiarito successivamente da un esempio.

³ La descrizione ricalca in gran parte quella contenuta in [TBL1997]

⁴ **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.).

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:

- possono essere *memorizzati come dati*, in una risorsa, che può quindi contenere informazioni relative a se stessa o ad un' altra risorsa;
- possono essere *descritti da altri metadati*, e così via.

2.2 - Il Resource Description Framework

Automatizzare il Web restando ancorati alla sua architettura originaria, in cui tutte le informazioni erano *machine-readable*, ma non *machine-understandable*, era un obiettivo molto difficilmente raggiungibile, mentre i *metadati* sembrano offrire una soluzione al problema. 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.

Resource Description Framework (RDF) 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. Può essere utilizzato e portare vantaggi sono in molti settori, quali, ad 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..

RDF 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 ([RDFMSS], [RDFSS]):

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

2.3 - RDF Data Model

RDF fornisce un modello per descrivere le risorse, che 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.

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

RDF permette di descrivere anche fatti complessi. Per esempio, il fatto espresso dalla concatenazione delle 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*

verrebbe rappresentato dal diagramma di Figura 2

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. I limiti pratici e logici per il numero di livelli, dipendono essenzialmente dalle caratteristiche e dalle tradizioni delle singole comunità che definiscono la descrizione delle risorse.

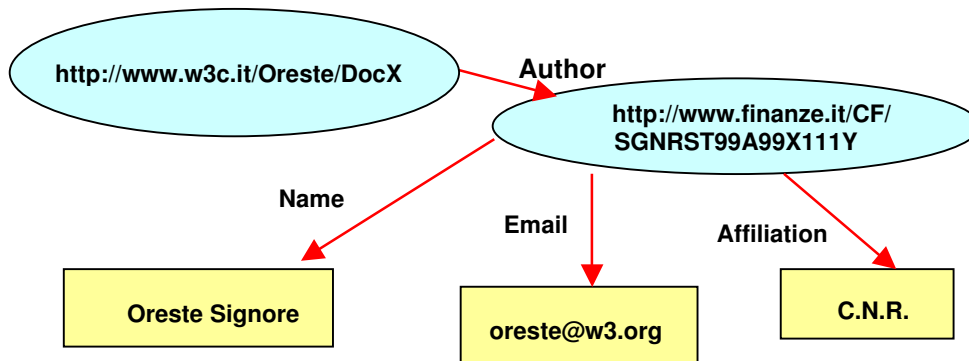


Figura 2 – La rappresentazione grafica di due statements RDF concatenati

2.4 - 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]) 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 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. 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.

3 - Trasformazione e presentazione dell' informazione

Molte sono ormai le applicazioni XML, ma, al di là delle soluzioni specifiche, va segnalata la possibilità di realizzare architetture applicative interessanti, distribuendo le funzioni tra client e server, o personalizzando la presentazione delle informazioni. La base su cui si fondano queste nuove applicazioni è la comprensione di un principio chiave dell' XML, cioè la *separazione tra contenuto e forma*. Ne scaturisce una architettura comune di riferimento, nella quale le informazioni, estratte dalla base dati aziendale, vengono strutturate in un documento XML, che viene successivamente trasformato nel formato più adatto per l' utente finale, mediante una trasformazione di stile. Come chiaramente descritto in [Holman2000], ogni volta che viene modificato lo stile della presentazione dei dati, si ha in realtà un doppio processo:

- in prima battuta, va *trasformata* l' informazione, dalla forma in cui è disponibile nell' organizzazione, a quella indicata per la distribuzione;
- successivamente questa informazione, ulteriormente ristrutturata, deve essere *presentata* nella maniera più efficace per il fruitore.

Nel flusso delle informazioni dall' origine alla fase di presentazione (*rendering*), è conveniente legare l' informazione alla sua forma finale il più tardi possibile,

mantenendo così la possibilità di utilizzare l' informazione anche per altri obiettivi durante l' intero processo. La pratica di modellare l' informazione sulla base di quello che sarà il suo utilizzo finale, per quanto molto diffusa, è da considerare poco accorta.

In generale, nella restituzione di un documento XML si hanno due passi distinti di trasformazione di stile:

- la *trasformazione* dell' istanza del vocabolario XML in una nuova istanza coerente con il vocabolario della semantica di restituzione
- la *formattazione* dell' istanza del vocabolario di restituzione nello user agent.

Per far fronte alle esigenze di questi due processi in modo indipendente, il Working Group del W3C responsabile dell' Extensible Stylesheet Language (XSL) ha messo a punto due Recommendation, una per la trasformazione dell' informazione, e l' altra per la restituzione (*rendering*).

La Recommendation *XSL Transformations* ([XSLT]) descrive un vocabolario riconosciuto da un XSLT processor per trasformare la struttura dell' informazione prodotta nell' ambito dell' organizzazione in un' altra struttura, adeguata per i successivi passi elaborativi.

La Recommendation *Extensible Stylesheet Language* ([XSL]) descrive un vocabolario riconosciuto dall' agente di rendering per trasformare i formati espressi in forma astratta in particolari mezzi di presentazione.

È importante tener presente che XSLT è stato progettato essenzialmente per *trasformare i vocabolari XML nel vocabolario XSL di formattazione*, per cui sono assenti alcune funzionalità che lo rendano uno strumento di uso affatto generale, anche se è utilizzabile per la maggior parte delle esigenze di trasformazione. Per questo motivo, i fogli di stile XSLT vengono spesso utilizzati in contesti diversi da quelli legati alla restituzione (*stylesheet rendering*) e sono detti *transformation script*.

Un esempio di trasformazione⁵ è nella sequenza di Figura 3, Figura 4, Figura 5, Figura 6.

```
<?xml version="1.0" ?>
<rubrica>
  <persona>
    <nome>Cristian</nome>
    <cognome>Lucchesi</cognome>
    <telefono>2116</telefono>
    <indirizzo>
      <via>Via Moruzzi, 25</via>
      <cap>56100</cap>
      <citta>Pisa</citta>
    </indirizzo>
  </persona>
  <persona>
    <nome>Oreste</nome>
    <cognome>Signore</cognome>
    <telefono>2995</telefono>
  </persona>
</rubrica>
```

Figura 3 – La rubrica telefonica in XML

⁵ Ringrazio il collega e amico Cristian Lucchesi per aver predisposto il materiale dell' esempio.

Tutti gli input devono essere documenti XML well formed, quindi non è possibile elaborare file HTML che seguono convenzioni lessicali non XML, ma è possibile elaborare file XHTML (*Extensible Hypertext Markup Language*)⁶

```
<?xml version="1.0" ?>
<xsl:stylesheet
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
  <xsl:output doctype-system=
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"
    doctype-public="-//W3C//DTD XHTML 1.0 Transitional//EN"
    indent="yes" media-type="text/html" method="html" />
  <xsl:template match="rubrica">
    <html xml:lang="it" lang="it">
      <head>
        <title>Rubrica di Esempio in HTML</title>
        <meta http-equiv="Content-Type"
          content="text/html; charset=iso-8859-1" />
      </head>
      <body>
        <h1>Rubrica Telefonica</h1>
        <table border="1">
          <tr>
            <td>Nome</td>
            <td>Cognome</td>
            <td>Telefono</td>
          </tr>
          <xsl:apply-templates />
        </table>
      </body>
    </html>
  </xsl:template>
  <xsl:template match="persona">
    <tr>
      <td>
        <xsl:value-of select="nome" />
      </td>
      <td>
        <xsl:value-of select="cognome" />
      </td>
      <td>
        <xsl:value-of select="telefono" />
      </td>
    </tr>
  </xsl:template>
</xsl:stylesheet>
```

Figura 4 – Lo script di trasformazione

In una architettura a tre livelli, è possibile inviare informazione strutturata al web browser utilizzando XSLT sull' *host*, sullo *user agent*, o su *entrambi*. Da un punto di vista tecnico, il server può distribuire il carico elaborativo ad agenti XML/XSLT inviando una combinazione di stylesheet e informazione sorgente da trasformare sul lato client; oppure può effettuare tutte le trasformazioni direttamente sul server, in

⁶ File XHTML possono essere generati da file HTML utilizzando un tool disponibile gratuitamente sul sito W3C (<http://www.w3.org/People/Raggett/tidy/>), che corregge le codifiche HTML errate e segnala quelle che non riesce a correggere.

modo da poter supportare anche user agent che supportano solo vocabolari HTML o HTML/CSS, o WML ([Lee2000]).

```
<?xml version="1.0"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xml:lang="it" lang="it">
  <head>
    <title>Rubrica di Esempio in HTML</title>
    <meta http-equiv="Content-Type"
      content="text/html; charset=iso-8859-1" />
  </head>
  <body>
    <h1>Rubrica Telefonica</h1>
    <table border="1">
      <tr>
        <td>Nome</td>
        <td>Cognome</td>
        <td>Telefono</td>
      </tr>
      <tr>
        <td>Cristian</td>
        <td>Lucchesi</td>
        <td>2116</td>
      </tr>
      <tr>
        <td>Oreste</td>
        <td>Signore</td>
        <td>2995</td>
      </tr>
    </table>
  </body>
</html>
```

Figura 5 – La rubrica telefonica in XHTML



Figura 6 – La rubrica telefonica nella presentazione all'utente finale

Indipendentemente dalle caratteristiche del browser, la scelta se trasferire documenti con markup XML o trasformarli in documenti HTML (semanticamente

meno ricchi) può dipendere da considerazioni legate all' opportunità di rendere disponibili a tutti le informazioni contenute nel documento XML completo. Una scelta potrebbe essere quella di trasformare il contenuto del documento in base alle caratteristiche del potenziale fruitore. Infine, per diminuire l' utilizzo della banda di trasmissione e incrementare l' elaborazione distribuita, è anche possibile utilizzare due processi XSLT, uno per trasformare il markup delle informazioni in un generico formato di distribuzione, e un altro sul client agent, per la personalizzazione,

Va comunque tenuto presente che esistono due linee di pensiero: secondo alcuni è necessario porre un "*semantic firewall*" tra i dati aziendali e gli utilizzatori, che quindi potrebbero anche essere dei semplici browser HTML, destinati solo alla presentazione di informazioni spogliate di qualunque semantica. Altri invece ritengono il Web debba comunque evolvere verso un *semantic web* (di cui si parla nel seguito) in cui l' informazione trasmessa debba sempre essere semanticamente ricca, ed essere poi elaborata da user agent intelligenti e dal comportamento rispettoso delle regole.

4 - Un aspetto spesso dimenticato: l' accessibilità dei siti Web

Non è possibile non menzionare la *Web accessibility*⁷ importante per vari motivi:

- l' uso del Web si va diffondendo rapidamente in tutta la società, ed è la tecnologia che nella storia ha avuto la diffusione più rapida;
- il Web costituisce un ostacolo per varie forme di disabilità
- milioni di persone hanno difficoltà nell' accesso al Web;
- alcuni siti Web **devono** essere accessibili (molti Governi, tra cui quello Italiano, hanno emanato linee guida per garantire l' accessibilità dei siti).

La Web accessibility ha una valenza *sociale*, ma anche *economica* (costituisce un mercato rilevante) e *tecnologica*, dato che la progettazione che tiene conto dei potenziali handicap porta dei benefici a tutti gli utenti, quando si trovano in condizioni ambientali difficili (dispositivi mobili, eccessiva illuminazione, elevato rumore di fondo, banda limitata, mani e occhi impegnati).

La *Web Accessibility Initiative (WAI)* del W3C ha operato in modo efficace per assicurare che le tecnologie Web supportino l' accessibilità, il cui supporto è già incluso in alcune specifiche, e ha sviluppato alcune Guideline che giocano un ruolo critico nel rendere accessibile il Web:

- *Web Content Accessibility Guidelines 1.0* ([WCAG]) che spiega agli autori come creare contenuti Web accessibili.
- *Authoring Tool Accessibility Guidelines 1.0* ([ATAG]) che spiega agli sviluppatori come progettare authoring tool utilizzabili dai portatori di handicap;
- *User Agent Accessibility Guidelines 1.0*, ([UAAG]) che illustra il ruolo degli user agent nel soddisfare le esigenze dei disabili.

⁷ "*The power of the Web is in its universality. Access by everyone regardless of disability is an essential aspect.*" -- Tim Berners-Lee, Direttore del W3C Director e inventore del World Wide Web.

5 - Il Semantic Web

Nella visione di Berners-Lee, il Web ha una architettura a livelli (Figura 7), che verrà sviluppata completamente nel giro di vari anni.

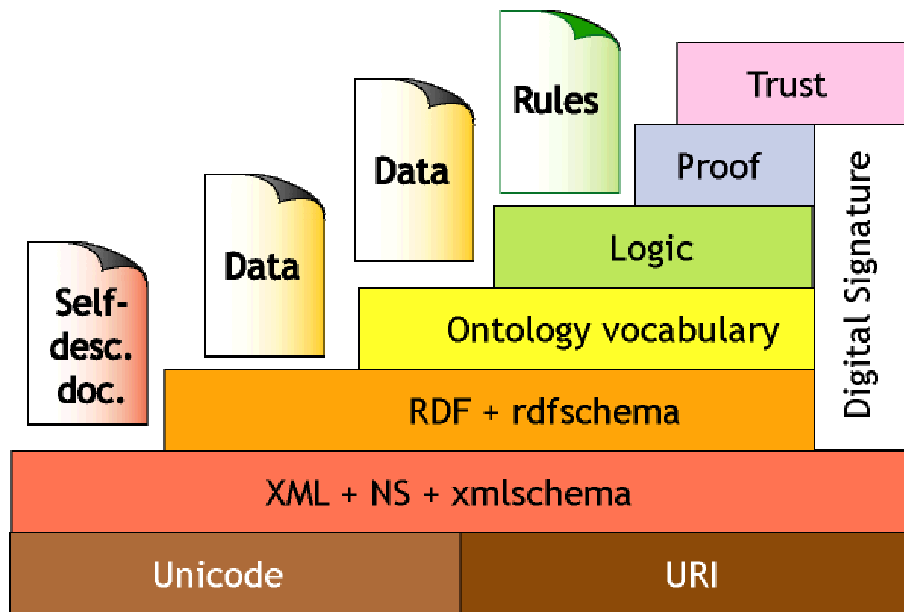


Figura 7 - L'architettura del Semantic Web
(<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>)

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.

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.

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.

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.

L'intera comunità scientifica sta investendo molte energie nel settore del 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. Le applicazioni Web e le tecnologie relative crescono con un ritmo frenetico e spesso disordinato, e potrebbe essere difficile operare le scelte strategiche più giuste. Senza alcun dubbio, l'utilizzo di XML si va diffondendo con una rapidità e una pervasività che testimoniano come molti abbiano compreso che si tratta di una famiglia di tecnologie centrale nello sviluppo del Web.

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.

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]).

Bibliografia

- | | |
|--------------|---|
| [ATAG] | <i>Authoring Tool Accessibility Guidelines 1.0</i> ,
http://www.w3.org/TR/WAI-AUTOOLS/ |
| [DC] | The Dublin Core Home Page, URL: http://dublincore.org/ |
| [Holman2000] | Holman G. Ken: <i>What is XSLT? (I)</i>
http://www.xml.com/lpt/a/2000/08/holman/s1.html |
| [IRDF] | Introduction to RDF Metadata, W3C NOTE 1997-11-13, Ora Lassila,
URL: http://www.w3.org/TR/NOTE-rdf-simple-intro |
| [Lee2000] | Lee Wei Meng: <i>Transforming XML into WML</i> ,
http://www.wirelessdevnet.com/channels/wap/training/xslt_wml.html |

- [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]
[RDFMSS] <http://www.sfc.keio.ac.jp/~ted/>
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/>
- [Signore2001] Signore O.: *Il ruolo centrale di XML nell' evoluzione del Web*, XML Day Milan, Conference proceedings, Milan, September 21
- [TBL1997] Tim Berners-Lee: *Metadata architecture*, (1997),
<http://www.w3.org/DesignIssues/Metadata.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>
- [UAAG] *User Agent Accessibility Guidelines 1.0*, <http://www.w3.org/TR/WAI-USERAGENT/>
- [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>
- [WCAG] *Web Content Accessibility Guidelines 1.0*, <http://www.w3.org/TR/WAI-WEBCONTENT/>
- [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/>
- [XMLschema0] XML Schema Part 0: Primer - W3C Recommendation - 2 May 2001
<http://www.w3.org/TR/xmlschema-0/>
- [XMLschema1] XML Schema Part 1: Structures - W3C Recommendation - 2 May 2001 <http://www.w3.org/TR/xmlschema-1/>
- [XMLschema2] XML Schema Part 2: Datatypes - W3C Recommendation - 2 May 2001
<http://www.w3.org/TR/xmlschema-2/>
- [XMLstylesheet] Associating Style Sheets with XML documents Version 1.0 - W3C Recommendation - 29 June 1999, <http://www.w3.org/TR/xml-stylesheet/>

[XSL]	<i>Extensible Stylesheet Language (XSL)</i> , Version 1.0 W3C Recommendation 15 October 2001 http://www.w3.org/TR/2001/REC-xsl-20011015/xslspecRX.pdf
[XSLT]	<i>XSL Transformations (XSLT)</i> Version 1.0, W3C Recommendation 16 November 1999, http://www.w3.org/TR/xslt

Appendice: alcune informazioni sul W3C

Le W3C Recommendation

Le W3C Recommendation sono il risultato di un processo cooperativo, che prevede una serie di passi e di documenti prodotti. Alcuni documenti sono riservati ai partecipanti ai gruppi di lavoro, altri sono disponibili per i membri, che votano per approvarli o modificarli, altri sono pubblici.

Anche se *Network Computing* cita il W3C nella sua lista dei "*Ten Most Significant Standards Groups*", dal punto di vista formale il W3C non è un organo di standardizzazione. Tuttavia, va riconosciuto che il W3C è una comunità di membri che cooperano spontaneamente per definire le linee guida e le specifiche, verificando che esse siano realmente implementabili, e mantiene stretti contatti con gli organi di standardizzazione e con gli User Forum.

La partecipazione al W3C

I membri del W3C, ad oggi, sono più di 500 ([W3CML]), e comprendono organizzazioni di vario tipo, che collaborano allo sviluppo del Web⁸. I vantaggi dell'associazione al W3C sono riconducibili a un guadagno in termini di maggiore informazione, miglioramento di immagine, coinvolgimento attivo e possibile influenza nell'evoluzione del Web, maggiore prontezza nel seguire l'evoluzione del mercato. Il punto essenziale dell'associazione al W3C consiste proprio nella partecipazione attiva allo sviluppo del Web, assicurando:

- Maggiore conoscenza dei processi evolutivi e possibilità di partecipare ai processi decisionali
- Conoscenza anticipata dell'evoluzione delle raccomandazioni, e quindi opportunità per svolgere attività tecnologica di punta, sviluppando prototipi in fase con lo sviluppo tecnologico e realizzando prima degli altri applicazioni e prodotti conformi alle Recommendation.

Per migliorare la sua presenza e i rapporti con le singole comunità nazionali, il W3C ha creato un certo numero di entità locali, denominate W3C Offices (<http://www.w3.org/Consortium/Offices/>).

L'Ufficio Italiano W3C (<http://www.w3c.it>), che ha iniziato la sua attività nel 1999, ha dato vita ad una iniziativa denominata *WebLab*, punto di incontro tra le attività di ricerca e l'integrazione e sviluppo di tecnologie Web.

⁸ Al momento, i membri italiani sono undici: C.N.R., CSP di Torino, Presidenza del Consiglio dei Ministri, Telecom Italia, Università di Bologna, CINECA, Dipartimento di Informatica dell'Università di Pisa, POSTECOM, Università Commerciale "Luigi Bocconi", Omnitel-Vodafone, CSI Piemonte