

Web semantico e ontologie

Daniela Canali

Istituto tecnologie biomediche
CNR - Roma
daniela.canali@tin.it

Nuovi sviluppi della comunicazione in rete

La crescita del web e il fenomeno della profondità e difficoltà di rintracciabilità delle risorse pone l'attenzione sui fondamenti semantici. L'emergere delle ontologie come oggettivazione di *significati intesi* e di mediazione semantica in domini specifici e specialistici assume una dimensione guida per i nuovi sviluppi della comunicazione web. In passato la ricerca sulle ontologie è stata piuttosto limitata alla sfera filosofica, ma ora ha assunto un ruolo specifico in diversi campi della ricerca come la rappresentazione e l'ingegneria della conoscenza, la creazione di modelli qualitativi, la progettazione di database e sistemi informativi, il reperimento e l'estrazione delle informazioni, lo sviluppo di software orientati agli oggetti, la gestione e l'organizzazione della conoscenza, lo sviluppo di sistemi basati su agenti intelligenti.

La parola ontologia genera spesso confusione a causa delle numerose interpretazioni dipendenti dal contesto:¹ è infatti usata per descrivere realtà con livelli strutturali diversi, quali le tassonomie (come la gerarchia di Yahoo)² e gli schemi di metadati (come il Dublin Core).

Alla base del web semantico³ ci sono ontologie con un livello strutturale elevato ed esse devono descrivere specificamente i concetti di *classe* (cose generiche nei vari domini di interesse), di *relazioni* esistenti tra le cose e di *proprietà* (o attributi) di tali cose.

In ambito informativo per ontologie intendiamo insieme di modelli concettuali (definiti anche *strutture di metadati*)⁴ costituiti da accordi sui concetti e sulle relazioni tra di essi, rappresentativi di un dominio di conoscenza: volendo utilizzare una definizione sintetica potremmo dire che un'ontologia è la specifica di una concettualizzazione.⁵

In ambito digitale le ontologie vengono normalmente costruite a partire dalle categorizzazioni del mondo reale e in particolare dei prodotti a stampa, sebbene tale derivazione diretta spesso si riveli inadeguata al contenuto digitale. Prendere a modello i supporti cartacei non soddisfa infatti le estese possibilità del contenuto digitale, come ad esempio l'aumento del contenuto originale con altre informazioni relative o la riformattazione del contenuto basata sull'interazione con l'utente.

Prima di costruire un'ontologia bisogna avere ben chiaro il significato definito e come esso può essere applicato ai metadati semantici. La natura concettuale e contestuale del significato è importante: secondo alcuni linguisti, gli elementi del linguaggio hanno una gamma di interpretazioni e la corretta interpretazione deriva dagli elementi circostanti il linguaggio, idea supportata nella *teoria del significato* proposta da Wittgenstein, secondo cui il significato di una parola è il suo uso nel linguaggio. Sebbene sia impossibile ottenere un accordo e un'interpretazione comune su tutto il mondo reale, è però possibile catturare la semantica essenziale del contenuto attraverso la struttura e la definizione di importanti caratteristiche dell'essenza del contenuto in un'ontologia, limitando le descrizioni a un certo dominio e ad un determinato livello di dettaglio.

La costruzione di un'ontologia è laboriosa e richiede la cooperazione di esperti del dominio, soprattutto per costruirne un buon livello di astrazione che ne garantisca la *stabilità* (le informazioni dettagliate possono velocemente perdere valore a causa dei cambiamenti ad un certo livello di dettaglio), l'*accessibilità*, la *natura schematica* e l'*eliminazione di dettagli irrilevanti* (il problema con le

¹ GIACOMO DEVOTO – GIAN CARLO OLI, *Il Dizionario della lingua italiana*, Firenze, Le Monnier, 2002 riporta per *Ontologia* la seguente definizione: "La dottrina filosofica relativa ai caratteri universali dell'ente, corrispondente alla 'prima filosofia' del più maturo Aristotele, detta poi 'metafisica': è tradizionalmente considerata il fondamento di ogni sistema oggettivistico" e di *Ontologico*: "Riguardante la natura e la conoscenza dell'essere come oggetto in sé; talvolta (contrapposto a soggettivo), improntato a un senso metafisico e oggettivo del reale".

² <<http://www.yahoo.com>>.

³ "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" (TIM BERNERS-LEE – HENDLER JAMES – ORA LASSILA, *The Semantic Web. A new form of web content that is meaningful to computers will unleash a revolution of new possibilities*, Scientific American, May 2001, <<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>>).

⁴ SAMI JOKELA, *Metadata enhanced content management in media companies*, dissertation for the degree of Doctor of technology, Helsinki University of Technology Department of Computer Science and Engineering Software Business and Engineering Institute, 2001, <<http://lib.hut.fi/Diss/2001/isbn9512256932/isbn9512256932.pdf>>.

⁵ TOM GRUBER, *What is an ontology?*, <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>.

astrazioni è che esse rappresentano una porzione della realtà da cui derivano e possono pertanto essere riduttive). Le ontologie vengono sviluppate per formalizzare e modellare le caratteristiche semantiche essenziali di un dominio di conoscenza e, per evitare qualsiasi ambiguità, durante lo sviluppo sono esplicitati, rappresentati e descritti le ipotesi chiave, il vocabolario e i principi del dominio. È importante notare la natura iterativa dello sviluppo di un'ontologia e la necessità di riflettere i cambiamenti nell'ambiente di riferimento: perché essa diventi veramente utile è necessaria una revisione costante e inoltre, anche quando si riesce a produrre un'ontologia che concettualizza in modo soddisfacente un certo dominio, è questo inevitabilmente a cambiare: sia che emergano nuovi concetti o si verifichino cambiamenti semantici di vecchi concetti, tutto deve riflettersi nell'ontologia che altrimenti rischia un graduale deterioramento.

La costruzione di ontologie dovrebbe concentrarsi solo sulle qualità rilevanti per l'essenza del contenuto e valide per l'uso inteso, tenendo in conto il numero degli utenti, quanto valore i metadati producono per questi utenti e i costi relativi: se questi sono superiori al valore prodotto non c'è ragione di realizzare un'ontologia, anche se in effetti alcune qualità che valorizzano l'essenza del contenuto sono difficili da identificare. Per esempio per molti utenti è importante l'attualità delle informazioni ma la distribuzione richiede metadati e ciò implica ulteriori ritardi nella produzione di contenuto. Una soluzione potrebbe essere quella di produrre i metadati in due fasi: fornendo inizialmente solo un set minimo per la gestione della disseminazione e successivamente aggiungere altri metadati semantici.

La modellizzazione di un dominio è condizionata da tre elementi:

- *information feeds* (capire non solo quali tipi di contenuto esistono o saranno disponibili ma anche quale tipo di informazione, formato e altre caratteristiche esso ha);
- *user needs* (perché gli utenti vogliono il contenuto e capire quali sue caratteristiche sono importanti);
- *intended use* (sapere quali caratteristiche sono necessarie nella produzione, nel recupero e nell'uso del contenuto, non solo per le applicazioni e i servizi esistenti e pianificati ma anche per tutte le attività relative al contenuto nella catena del valore).

Se l'ontologia non è direttamente utilizzata dall'uomo può avere una struttura molto complessa:⁶ quindi è importante capire chi sono gli utilizzatori e se l'ontologia sarà visibile all'amministratore del sistema e/o all'utente finale, e se invece è nascosta e usata solo dai computer.⁷

L'automazione è una componente importante delle onto-

logie: se la creazione dei metadati è automatica, il processo può essere molto dettagliato e portare a metadati semantici complessi. L'estrazione di metadati completamente automatizzata, senza alcun controllo umano, non è raccomandabile, poiché si tratta di un processo molto complicato che può determinare una qualità inferiore dei metadati. Il lavoro di routine nella creazione dei metadati semantici dovrebbe essere quanto più possibile automatizzato ma non privo del controllo di esperti.

Anche se l'ontologia viene intesa come assolutamente ideale in un certo momento nel tempo, essa è inevitabilmente soggetta a degrado e richiede mantenimento. Alcuni domini sono più statici di altri: le news rappresentano ad esempio un dominio molto dinamico, con molte tematiche e molti utenti con le loro diverse interpretazioni dei concetti, mentre alcuni domini scientifici, come la chimica, in cui l'informazione è spesso basata su fondamenti ben definiti come le tavole periodiche, sono più statici. In questi domini i nuovi contenuti possono essere descritti nelle strutture ontologiche esistenti apportando al limite lievi cambiamenti.

Le ontologie definiscono dunque i termini usati per descrivere e rappresentare un'area della conoscenza e sono utilizzate da umani e da database e applicazioni per condividere informazioni in un particolare dominio, ossia un settore tematico specifico: esse includono le definizioni utilizzabili dai computer, sulla base di concetti base del dominio, e le relazioni tra di essi, codificando la conoscenza e rendendola riutilizzabile.

Attualmente le aree di applicazione sono disparate, dal commercio elettronico alle biblioteche digitali, alla medicina, ai sistemi informativi, e il ruolo chiave delle ontologie nello scambio di dati basato sui contenuti in queste aree è attestato dall'interesse mostrato da molti organismi e iniziative di standardizzazione internazionali, tra cui ISO, ANSI, il W3C,⁸ IEEE ecc.

Nel campo emergente del Semantic Web le ontologie sono un modo di rappresentare la semantica dei documenti e renderla utilizzabile da applicazioni web e da agenti intelligenti.

Mappatura di ontologie

All'interno di una comunità le ontologie possono rivelarsi molto utili per strutturare e definire il significato dei termini dei metadati usati.⁹ Esse si rivelano però critiche per applicazioni che cercano o mescolano informazioni da comunità diverse: uno stesso termine può essere usato con significati diversi in contesti differenti e termini diversi

⁶ I linguisti sono giunti alla conclusione che le gerarchie tassonomiche raramente hanno più di cinque o sei livelli, e anche questo numero è poco comune. Comunque tali parametri non sono necessariamente applicabili ai vocabolari tecnici.

⁷ Una categorizzazione come Yahoo, per esempio, è progettata per mantenere un certo numero di scelte parallele: se il numero delle scelte aumenta oltre un certo limite, la categorizzazione viene sfrondata.

⁸ <<http://www.w3.org/TR/webont-req/#onto-def>>.

⁹ Nell'ambito degli editor automatici di metadati va detto che la loro considerazione delle ontologie può semplificare notevolmente la creazione di metadati per risorse disponibili nel web.

possono essere usati per documenti che hanno lo stesso significato. Il processo di mappatura consente di rendere compatibili contenuti diversi e i loro metadati semantici, definendo regole con cui si possono omogeneizzare i metadati forniti da fonti eterogenee di dati. La mappatura può avere livelli diversi di complessità e può coinvolgere la traduzione di un'ontologia nella sua controparte semanticamente equivalente in altre ontologie. Un'ontologia di origine può usare ad esempio il concetto *car* mentre quella di destinazione contiene il concetto *automobile*: in questo caso il conflitto si risolve rinominando il primo concetto. Ma quando ontologie di origine e di destinazione non contengono concetti semanticamente equivalenti sorgono problemi molto più complessi, come l'uso di concetti multipli o cambiamenti nell'ontologia originaria. È comunque preferibile passare da un formato a un altro piuttosto che produrre formati intermedi: più sono i formati, maggiore è il vantaggio di usare metodi riutilizzabili e un comune formato ontologico di metadati semantici originario cui tutte le parti rilevanti devono aderire, anche se è normale che diversi stadi del contenuto abbiano proprie ontologie.

Resource Description Framework e il relativo schema¹⁰ affrontano questo problema permettendo di associare una semantica semplice agli identificatori. Con lo schema RDF è possibile infatti definire classi che possono avere sottoclassi e sovraclassi multiple, e possono definire proprietà che hanno a loro volta sottoproprietà, domini e *range* di valori.

In questo senso lo schema RDF rappresenta un linguaggio di ontologia semplice, ma per ottenere interoperabilità tra numerosi schemi, sviluppati e gestiti autonomamente, è necessaria una semantica più ricca. Il grande sforzo attual-

mente consiste nel definire cosa è necessario per costruire un linguaggio di ontologia per il web, considerata la natura di ambiente unico della rete.¹¹

Il modello di metadati ABC è stato sviluppato nell'ambito del progetto internazionale di *digital library* Harmony¹² per fornire un modello concettuale comune che faciliti l'interoperabilità tra ontologie di metadati appartenenti a domini diversi. Sulla base di tale modello, che consta di categorie (temporalità, attualità e astrazione), di classi e proprietà è stato costruito un deposito di descrizioni RDF e un'interfaccia di ricerca dotata delle più sofisticate modalità di interrogazione. Specificamente progettate per modellare la creazione, l'evoluzione e la mobilità degli oggetti nel tempo, sono molto interessanti le possibili applicazioni del modello nell'ambito dei metadati di gestione di ambiente multimediale nei processi documentali.

Ontologie e portali

Le ontologie possono essere utilizzate per delineare nuovi usi del web e per migliorare le applicazioni web esistenti, quali ad esempio i portali, ossia siti web che oltre a fornire contenuto informativo su un determinato argomento¹³ offrono alti livelli di interattività e servizi ormai consueti quali la possibilità di ricevere news, di comunicare con altri membri della comunità, di costruire comunità, di trovare link ad altre risorse web di interesse comune. Il successo di un portale è legato alla sua capacità di rappresentare un valido punto di partenza per l'individuazione di contenuto interessante, sempre più di frequente oggi immesso dai membri stessi della comunità.¹⁴ Gli stessi fornitori di dati

¹⁰ RDF (Resource Description Framework) è una specifica del World Wide Web Consortium (W3C) rilasciata nel 1999. Progettato per l'elaborazione di metadati web, RDF fornisce una piattaforma di interoperabilità tra applicazioni sul web che si scambiano informazioni comprensibili dalle macchine. Tale modello consente di rappresentare i metadati in una sintassi per la codifica e la trasmissione secondo modalità che favoriscono l'interoperabilità tra server e client web sviluppati in modo indipendente. Il framework facilita l'elaborazione automatica delle informazioni e può essere utilizzato in diverse aree di applicazione: nella ricerca di risorse (*resource discovery*), per migliorare le capacità dei motori di ricerca, per descrivere il contenuto e le relazioni disponibili in un particolare sito web, pagina o digital library, per facilitare condivisione e scambio di conoscenza tramite gli agenti intelligenti, nella valutazione del contenuto, nella descrizione di collezioni di pagine che rappresentano un singolo documento *logico*, nella descrizione del copyright di pagine web, per esprimere le preferenze di un utente, per la gestione della privacy di un sito.

¹¹ NICOLA GUARINO, *Formal ontology and information systems*, "Formal Ontology in information systems", proceedings of the 1st International Conference, Trento, Italy, 6-8 June, IOS Press, 1998, <<http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/FOIS98.pdf>>.

¹² Il progetto internazionale Harmony è finanziato dal Distributed Systems Technology Centre (DSTC, Australia), dal Joint Information Systems Committee (JISC, UK) e dalla National Science Foundation (USA) con l'obiettivo di definire metodi e modelli per la descrizione della varietà di contenuto ricco (*rich content*) che sempre di più popola il web e le biblioteche digitali. Il sito è ricchissimo di materiale informativo: <<http://metadata.net/harmony/>>. Per una descrizione dettagliata di veda CARL LAGOZE – JANE HUNTER, *The ABC ontology and model*, "Journal of Digital Information", 2, 2001, <<http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/lagoze-final.pdf>>.

¹³ Un portale è sostanzialmente un aggregatore di informazione che offre un servizio di navigazione sul WWW facilitando il lavoro di ricerca: nati come evoluzione dei motori di ricerca, i portali hanno associato agli strumenti tipici di questi (*search engines* e categorizzazione delle informazioni) altri servizi, informativi e non, allo scopo di proporsi come accesso preferenziale e guida per la navigazione via Internet. Negli ultimi tempi sono state individuate due tendenze di sviluppo: da un lato si ha uno sforzo verso la sempre maggiore integrazione dei servizi di navigazione, dall'altro si assiste a un passaggio da un'ottica di offerta di contenuti di stampo generalista a una di posizionamento di dominanza in un settore di nicchia, con obbligo quindi di una notevole specializzazione. In particolare questi portali tematici vengono definiti *vortali*, portali verticali, detti anche *portali di secondo livello*. Vedi anche TAMAR SADEH – JENNY WALZER, *Library portals: toward the Semantic Web*, New Library, 104 (2003), 1184/5, <http://www.exlibris.co.il/meta-lib/Library_portals_toward_the_semantic_Web.pdf>.

¹⁴ GRIGORIS KARVOUNARAKIS – VASSILIS CHRISTOPHIDES – DIMITRIS PLEXOUSAKIS, *RQL: a declarative query language for RDF*, "D-Lib Magazine", 7 (2001), 12, <www.dlib.org/dlib/december01/12inbrief.html#KARVOUNARAKIS>. Una tecnologia attualmente molto diffusa di immis-

etichettano il contenuto con informazioni che prendono la forma di metatag identificative dell'argomento del contenuto. Un semplice indice per soggetto può rivelarsi insufficiente a ottimizzare il recupero dei contenuti: il portale che implementa un'ontologia per la comunità di riferimento, fornendo la terminologia per descrivere il contenuto e gli assiomi che definiscono i termini, consente un controllo più intelligente.

Un'ontologia può ad esempio includere termini come "giornale", "pubblicazione", "persona", "autore" e definizioni che asseriscono che "tutti i giornali sono pubblicazioni o tutte le persone sono autori". Combinate con i fatti, queste definizioni permettono di inferire altri fatti che sono necessariamente veri. Queste inferenze, a loro volta, possono permettere agli utenti di ottenere dal portale dei risultati di ricerca che è impossibile ottenere con i tradizionali sistemi di recupero dell'informazione.

Una simile tecnica richiede necessariamente che i *content providers* annotino le loro pagine con il linguaggio dell'ontologia, fatto che di solito si rivela piuttosto difficile da ottenere: la ricaduta che essi possono realizzare distribuendo il loro contenuto il più ampiamente possibile può essere un elemento a favore che potrebbe convincerli a fare questo lavoro di annotazione, del resto ormai possibile con modalità automatiche.

Un esempio di portale basato su un'ontologia è OntoWeb¹⁵ (Ontology-based Information Exchange for Knowledge Management and Electronic Commerce) presso la Free University of Amsterdam, un progetto cui partecipa anche l'Italia,¹⁶ che ha come scopo prioritario quello di raccordare i ricercatori e gli operatori dell'industria di varie aree di applicazione promuovendo il lavoro interdisciplinare e rafforzando l'influenza europea sugli sforzi di standardizzazione del Semantic Web.

Altro esempio di portale che usa la tecnologia Semantic Web e potrebbe trarre molto beneficio da un linguaggio ontologico è il progetto Open Directory,¹⁷ la grande directory del web costruita e mantenuta da una comunità vasta e globale di editori volontari. In questo caso è possibi-

le scaricare dal sito le copie RDF della base di dati della directory.

L'uso di ontologie nella navigazione di basi di dati è certamente interessante ma sarebbe molto riduttivo pensare a questo come al loro ruolo chiave.¹⁸ Le ontologie sono innanzi tutto degli schemi e come tali hanno lo scopo di organizzare un dominio in uno schema dei dati in grado di suddividerlo in tutte le classi di oggetti che hanno un ruolo nei suoi processi.

Come esempio di quanto detto si può portare il campo dell'*e-learning* in cui almeno un paio di grosse applicazioni sono state sviluppate secondo una struttura ontologica:¹⁹ organizzare una piattaforma secondo strutture ontologiche si rivela estremamente produttivo poiché ad esempio un corso di meccanica può contenere una parte relativa alla meccanica dei fluidi che può essere ripresa, e poi ampliata, in un corso di idraulica. Se i vari materiali didattici sono organizzati in unità (*learning objects*) ogni singola unità può essere collegata alle altre e ricomposta in un nuovo corso.

Ontologie e oggetti digitali non testuali

Le ontologie possono essere usate anche per fornire annotazioni semantiche a collezioni di immagini, oggetti audio, video o altri oggetti non testuali. Questo tipo di risorse sono di solito indicizzate tramite cattura o metatag, poiché è molto più complesso per le macchine estrarre una semantica significativa da oggetti multimediali rispetto a un testo in linguaggio naturale.

Poiché questi oggetti non testuali possono essere descritti in modo differente da parte di diversi indicizzatori, è importante che le possibilità di ricerca vadano oltre il semplice abbinamento di parole chiave: le ontologie possono catturare conoscenza aggiuntiva riguardo al dominio, da usare per migliorarne il recupero.

Le ontologie multimediali possono essere di due tipi: specifiche per supporto (possono avere tassonomie di diversi

sione di contenuti sul web da parte degli utenti è il *web log* (detto anche *blog*) ossia un sistema di messaggistica tramite web che evita il pesante e lungo lavoro di annotazione delle risorse. Tramite una procedura di immissione di dati in pochi campi predefiniti, viene inizializzata una descrizione base della risorsa eventualmente da integrare. Per un'esperienza diretta di questa procedura si rimanda al progetto europeo Hyperguide e al sito di fruizione Tool2know (T2Klog), <<http://194.183.21.37/hyperguide/Weblog.html>>, oltre che alla pubblicazione: DANIELA CANALI – AUGUSTA MARIA PACI – FILIPPO MARIA VINCIGUERRA, *A knowledge path to research, project and product: the experience of Hyperguide/Tool2know*, Roma, EUE (Editoria Università Elettronica), 2002.

¹⁵ <<http://starpc14.vub.ac.be:8000/OntoWeb/Browse/index.html>>.

¹⁶ I partecipanti italiani sono: OTH, LADSEB-CNR, Institute of Systems Science and Biomedical Engineering, Consiglio nazionale delle ricerche, CNR - ITBM, DSI - Università di Modena e Reggio Emilia.

¹⁷ <<http://dmoz.org/about.html>>.

¹⁸ La posizione di Nicola Guarino nella sua relazione alla tappa italiana del W3C Semantic Tour si rivela ancora molto legata al concetto di database e infatti delinea il ruolo delle ontologie nel Semantic Web come finalizzate all'interoperabilità semantica e all'information retrieval. L'ontologia è una concettualizzazione, una struttura formale della realtà così come è interpretata da un agente, indipendentemente dal vocabolario usato e dall'attuale occorrenza di una specifica situazione. La funzione dell'ontologia è dunque principalmente quella di catturare il *significato inteso* di un vocabolario secondo una concettualizzazione: si tratta di una teoria logica, i cui modelli catturano il più possibile i modelli *intesi*. Qual è dunque la differenza tra le ontologie e i vecchi modelli concettuali? In effetti, secondo Guarino, non troppa, poiché le ontologie rappresentano modelli concettuali più raffinati e con alcune attenzioni particolari (i modelli non sono accessibili *run time* e non hanno una semantica formale). Nel modello vengono specificati solo i vincoli importanti per le *queries*, mentre nelle ontologie si tiene conto dei vincoli necessari per il significato inteso.

¹⁹ <<http://edutella.jxta.org>> e <<http://www.merlot.org>>.

tipi di supporto e descrivono proprietà di supporti diversi)²⁰ e specifiche per contenuto (descrivono il soggetto di una risorsa). Questo secondo tipo, in particolare, non è specifico rispetto al supporto e ciò permette di riutilizzarle in altri documenti relativi allo stesso argomento. Tale riutilizzo potenzia la ricerca delle informazioni su un particolare soggetto, a prescindere dal formato della risorsa; se invece questo è rilevante, si possono combinare entrambi i tipi di ontologia.

Attualmente sono disponibili in rete applicazioni gratuite per la creazione di ontologie, come ad esempio Protégé-2000, sviluppato dalla Stanford Medical Informatics presso la Stanford University School of Medicine, che consente all'utente di costruire una propria ontologia di dominio, personalizzare il formato dei dati da immettere e inserire i dati.

Per meglio comprendere la struttura di un'ontologia possiamo ipotizzare la realizzazione di un'ontologia relativa a una rivista, premesso che la base di conoscenza (*knowledge base*) di un giornale può essere usata in molti modi diversi. Per il nostro esempio scegliamo:

- una lista di articoli pubblicati (indicando data e sezione di pubblicazione);
- le informazioni riguardo alle sezioni standard del giornale (sport, cronaca, cultura ecc);
- le informazioni sul personale;
- le informazioni relative alla pubblicità.

A partire da questa *knowledge base*, tra le molte applicazioni possibili quali sistemi di rappresentazione, si potrebbe costruire un sistema per il recupero, l'organizzazione e la risposta alle *queries* riguardo agli articoli pubblicati, un sistema per analizzare i redditi o la politica dei prezzi di annuncio, un sistema per la revisione dell'organizzazione per assicurare un equilibrio tra i redattori e gli editori e per garantire ad ogni sezione del giornale un editore responsabile.

Il Knowledge Systems Laboratory della Stanford University ha sviluppato Chimera,²¹ un software che supporta gli utenti nella creazione e nel mantenimento di ontologie distribuite sul web.

Nonostante la diffusione, la disponibilità in rete e l'illustrazione particolareggiata delle procedure di creazione, si tratta comunque di sistemi piuttosto complessi da utilizzare.

Quando la diffusione di ontologie nel web sarà massiccia, sarà necessario renderle interoperabili, definendo mappature tra i diversi schemi: questo problema è già affrontato da diversi progetti come lo Stanford Scalable Knowledge Composition (SKC) Project²² che ha sviluppato un nuovo approccio per risolvere il problema dell'eterogeneità semantica nei sistemi informativi, basandosi sull'algebra per comporre ontologie che rappresentano le terminologie appartenenti a domini distinti e autonomi.

Il Polo didattico e di ricerca di Crema ha costituito un gruppo di lavoro che studia e utilizza le ontologie per ottimizzare i processi di condivisione della conoscenza in vari campi applicativi, dai sistemi di produzione del software alla navigazione e reperimento di basi di documenti. Ha sviluppato Pocket Knowledge Hub, un pacchetto integrato di tool che permette di installare su semplice pc un piccolo sistema di Knowledge Management semantico:²³ attraverso un'ontologia definita dall'utente, il sistema consente di indicizzare documenti e di navigarvi in modo semantico, sia partendo da un albero gerarchico sia ricostruendo le relazioni di ogni singolo concetto.

Ontologie di dominio

Una ontologia di dominio descrive una specifica di categorie base definite dai concetti concreti e dalle relazioni che sorgono all'interno di una particolare applicazione. È quindi necessario prendere in considerazione i punti di vista di diversi esperti sui concetti e le relazioni del dominio come pure i vari obiettivi e i punti focali determinati dal contesto. Le ontologie di dominio hanno una bassa portabilità, poiché la possibilità di trasferirle in altre applicazioni è molto limitata. Il lavoro su di esse è però indispensabile per testare le ontologie generiche rilevanti e definire il framework ontologico di livello superiore, correlato sia al potere espressivo sia alla robustezza. Sono stati trovati metodi per specificare il grado di portabilità di concetti specifici di dominio, usando per esempio metodi descritti strettamente modulari.

Un esempio è lo sviluppo di una ontologia generica per il dominio delle prove cliniche nell'ambito del progetto OntoMed, costruito con il software OntoBuilder²⁴ tramite il sistema GOL-Clin che si basa sul General Ontological Language (GOL): in questo caso l'ontologia generica per prove cliniche Onto-Clin è stata aggiunta all'ontologia base General Formal Ontology (GFO). Nell'ambito di questo progetto viene sviluppato anche uno specifico Data Dictionary. OntoBuilder ha due componenti.

1) OntoExtract:

- estrae informazioni da fonti a testo libero non strutturato;
- produce ontologie leggere da testo naturale;
- memorizza e controlla da un server Sesame.

2) OntoWrapper:

- informazioni strutturate tabellari.

I punti di forza di OntoBuilder sono rappresentati dalla capacità di estrarre dai documenti testi in linguaggio naturale, dall'accuratezza in tale estrazione e dalla generazione di conoscenza.

Il progetto europeo OntoWeb riguarda lo scambio di in-

²⁰ Per esempio i video possono includere proprietà che identificano la durata e le interruzioni di scena.

²¹ <<http://www.ksl.stanford.edu/software/chimera/>>.

²² <<http://www-db.stanford.edu/SKC/>>.

²³ Basato su architettura Java e su un database Access, è facilmente installabile su qualsiasi macchina contenente una *java virtual machine*. Il pacchetto è scaricabile dal sito web: <<http://ra.crema.unimi.it/ontology/index.htm>>.

²⁴ <<http://www.onto-med.de/en/index.html>>.

Fig. 1 – Struttura multilivello del Semantic Web



formazioni basate su ontologie per la gestione della conoscenza e il commercio elettronico. Il web attuale è ancora un web prettamente sintattico, in cui la struttura del contenuto è disponibile ma il contenuto stesso è inaccessibile ai computer.

Nel discorso tenuto all'apertura della decima conferenza del W3C a Hong Kong, nel maggio 2001, Tim Berners-Lee sosteneva che nel prossimo futuro saranno possibili collegamenti telematici di qualunque genere e un'inedita forma del contenuto web – il Semantic Web – darà impulso a una vera e propria rivoluzione, ricca di possibilità.

Mentre la nostra mente organizza spontaneamente le informazioni sulla base di concetti, il web non funziona così e non c'è alcun modo di chiedere a Internet quali pagine web siano collegate a un *concetto* che ci interessa. Il Semantic Web rappresenta un'estensione del web attuale, in cui le informazioni sono strutturate con un senso compiuto, migliorando il lavoro tra le persone e le macchine, un web più *intelligente*, capace di gestire automaticamente le nostre richieste di informazione. Il termine Semantic Web, proposto proprio per la prima volta nel 2001 da Tim Berners-Lee, è associato all'idea di un web nel quale agiscano agenti intelligenti, ossia applicazioni in grado di

comprendere il significato dei testi presenti sulla rete e perciò di guidare l'utente direttamente verso l'informazione ricercata, oppure di sostituirsi a lui nello svolgimento di alcune operazioni.²⁵ 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.

Gli agenti informatici sono infatti in grado di comprendere le relazioni tra le informazioni e coordinano questa loro capacità di comprensione con le richieste specifiche dell'utente, collegando l'informazione presente nelle pagine web a concetti astratti organizzati in una gerarchia (l'ontologia), a sua volta descritta in un metadocumento.²⁶ L'adozione universale di standard aperti per la rappresentazione e la condivisione della metainformazione rende le macchine capaci di analizzare tutti i dati sul web e di capirne i contenuti:²⁷ l'infrastruttura tecnologica del web può ormai considerarsi consolidata mentre i problemi nascono dal punto di vista del contenuto.

L'idea del web semantico nasce quindi semplicemente estendendo l'idea di utilizzare schemi per descrivere domini di informazione in cui i metadati mappano i dati rispetto a classi, o concetti, di un certo schema di dominio, determinando strutture capaci di descrivere e automatizzare i collegamenti esistenti fra i dati. Il web semantico è multilivello, e si compone fondamentalmente di tre livelli: al più basso abbiamo i *dati*; i *metadati* riportano questi dati ai concetti di uno schema; nello *schema* (spesso chiamato *ontologia*) si esprimono le relazioni fra concetti che diventano classi di dati:

- l'*informazione* (le risorse web);
- i *metadati* (l'informazione relativa ai contenuti di una risorsa);
- le *ontologie* (la descrizione attraverso l'esplicitazione di una rete di relazioni del significato delle asserzioni usate relativamente alla descrizione dei dati). Vedi figura 1.

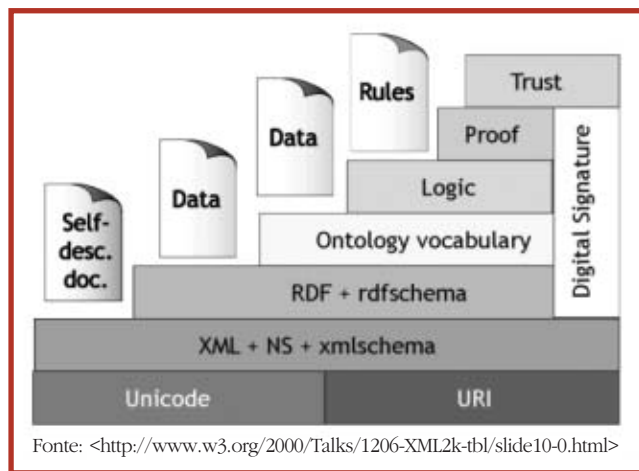
Un sistema così strutturato, basato sullo standard URI (Uniform Resource Identifiers) per la definizione univoca

²⁵ TIM BERNERS-LEE – JAMES HENDLER – ORA LASSILA, *The Semantic Web...*, cit.

²⁶ Ovviamente la gerarchia di concetti non è unica, ma ne esiste una diversa per ogni dominio applicativo di interesse, dalla medicina al controllo di gestione.

²⁷ TIME BERNERS-LEE, *L'architettura del nuovo web*, Milano, Feltrinelli, 2001. "Ho fatto un sogno riguardante il web... ed è un sogno diviso in due parti. Nella prima parte, il web diventa un mezzo di gran lunga più potente per favorire la collaborazione tra i popoli. Ho sempre immaginato lo spazio dell'informazione come una cosa cui abbiamo accesso immediato e intuitivo, non solo per navigare ma anche per creare.[...]". Nella seconda parte del sogno, la collaborazione si allarga ai computer. Le macchine diventano capaci di analizzare tutti i dati sul web, il contenuto, i link e le transazioni tra persone e computer. La "Rete Semantica" che dovrebbe renderlo possibile deve ancora nascere, ma quando l'avremo i meccanismi quotidiani di commercio, burocrazie e vita saranno gestiti da macchine che parleranno a macchine, lasciando che gli uomini pensino soltanto a fornire l'ispirazione e l'intuito.[...] Questo web comprensibile alle macchine si concretizzerà introducendo una serie di progressi tecnici e di adeguamenti sociali attualmente in fase di sviluppo".

Fig. 2 – Diagramma piramidale presentato da Tim Berners-Lee



di indirizzi Internet, su XML (Extensible Markup Language) e RDF (Resource Description Framework) nonché sull'uso delle ontologie, è in grado di riconoscere le risorse che contengono ciò che noi cerchiamo e garantisce l'accesso all'informazione, stabilendo la corrispondenza tra un segno e una sua certa semantica, e muovendosi in essa per reperire solo le referenze utili del segno, tralasciando quelle che non si accordano con la richiesta.

Nel modello di web molto più complesso, presentato da Tim Berners-Lee alla conferenza "Semantic Web on XML 2000", appare evidente, oltre al ruolo di base giocato da XML (con *Name Space* e *XMLschema*), la centralità di RDF e RDFSchema, che costituiscono il linguaggio per descrivere le risorse e i loro tipi.

Al di sopra di questo strato si pone il *livello ontologico*, in cui vengono descritte le relazioni tra i tipi di elementi senza però fornire informazioni su come utilizzare queste relazioni dal punto di vista computazionale. La firma digitale (*digital signature*), tecnica nota da qualche anno ma non ancora diffusa su larga scala, è di significativa importanza in diversi strati nel modello astratto del Semantic Web, poiché potrebbe essere utilizzata per stabilire la provenienza delle ontologie e delle deduzioni, oltre che dei dati. 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 (figura 2).

Il web semantico non si basa dunque su una rigida gerarchizzazione degli elementi e non è un database che divide tutti i dati in rigide categorie: le ontologie che definiscono le relazioni fra elementi possono sempre essere arricchite e specificate, pur essendo auspicabile che i vari domini utilizzino una stessa semantica.

²⁸ <<http://www.w3c.org>>. "Il web deve riuscire a cambiare lentamente, un passo alla volta, senza farsi bloccare o doversi riprogettare da zero. Ciò vale non solo per il web, ma anche per applicazioni, concetti, macchine e sistemi sociali su di esso edificati. Perché gli accessori muteranno ancora più alla svelta. Le applicazioni sul web non vengono create di colpo, nascono da una piccola idea per poi diventare più forti e complesse." Il concetto di evolvibilità era stato trattato da Tim Berners-Lee già nel 1998 e poi riproposto nel 2000. <<http://www.w3.org/DesignIssues/Evolution.html>>.

Fig. 3

```
<continent rdf:ID="Asia"/>
<rdf:Description rdf:ID="Asia">
  <rdf:type>
    <rdfs:Class rdf:about="#continent"/>
  </rdf:type>
</rdf:Description>
<rdf:Description rdf:ID="India">
  <is_part_of rdf:resource="#Asia"/>
</rdf:Description>
```

L'esempio nella figura 3 mostra come funziona la filosofia del Semantic Web, proponendo una porzione di codice RDF usata per descrivere l'India come parte del continente Asia: viene prima definito il continente "Asia", poi viene identificata l'India come parte del continente.

Fig. 4

```
<myOntology:Person>
  <myOntology:givenName>Dan</myOntology:givenName>
  <myOntology:isAuthorOf>
    <myOntology:Program />
  </myOntology:isAuthorOf>
</myOntology:Person>
```

La codifica riportata in figura 4 associa un programma all'autore, assegnando il nome di nascita "Dan" al programma (ad esempio un software online), specificando che ne è l'autore.

I collegamenti tra i concetti, che costituiscono la cosiddetta *rete semantica*, permettono a siti diversi di avere una propria definizione di un soggetto ricercato, perché il livello di inferenza consente alle macchine di ricollegare le varie definizioni, trascurando la necessità che due persone abbiano la medesima idea rigida di cosa "è" qualcosa. Come un bambino impara un'idea tramite contatti ripetuti, la rete semantica impara un concetto tramite contributi ripetuti da diverse fonti indipendenti, senza basarsi sui linguaggi naturali.²⁸ Un esempio è costituito dalle informazioni personali presenti in grande quantità nella rete, in calendari di conferenze, home page, elenchi telefonici ecc.: esse possono essere connesse e integrate dimostrando l'abilità di mescolare informazioni provenienti da fonti eterogenee. Attraverso questa rete di relazioni è possibile rispondere a domande tipo: qual è l'home page di questa persona?, qual è il numero di telefono del suo assistente?,avrò la possibilità di incontrarlo nell'arco di questo mese? e così via.

Tanto più le macchine sono in grado di leggere e comprendere il contenuto di una pagina web in maniera automatizzata, tanto più cresce il valore reale della rete semantica e tanto più è importante disporre di programmi in grado di collegare il contenuto web.

Pensando ad esempio alla redazione di materiali didattici si potrebbe immaginare un unico testo da cui generare una versione per studenti del primo anno, una per dottorandi, una per docenti, o addirittura che, in base a indicazioni date dal lettore all'inizio della consultazione (per esempio: "non ho ancora sostenuto tale esame" o "mi interessano soprattutto tali argomenti") si generi al momento il testo che verrà poi analizzato.

Leggendo le notizie sul web, mentre ora è necessario visitare singolarmente i siti dei vari giornali, le notizie disponibili nel web semantico possono essere raccolte in una pagina dove è possibile scegliere le fonti, il grado di aggiornamento degli articoli e altri parametri, ricavando in pratica una versione personale del giornale, con gli argomenti o gli autori di interesse.

Il web semantico definisce una convergenza di media (ipertesto, multimedia, realtà virtuale, reti neurali, agenti digitali) in cui si può teorizzare una visione simile a quella che Derrick de Kerckhove sosteneva per Internet, ossia di un ambiente vivo, di milioni di intelligenze umane perpetuamente al lavoro su qualcosa o su tutto con potenziale rilevanza per qualcuno e per tutti in una nuova condizione cognitiva che chiama *webness*.²⁹

Il valore centrale delle tecnologie della rete semantica, consiste nella capacità di presentare i contenuti esattamente nel formato che l'utente desidera e su qualsiasi dispositivo, che si tratti di un telefono cellulare, televisore, PDA (Personal Digital Assistance), personal computer o altro. In particolare l'XML affronta tre aspetti indispensabili per l'intelligenza mobile: "chi" (chi è l'utente),³⁰ "cosa" (che cosa vuole fare sul *web wireless*)³¹ e "dove" (dove si trova in un dato momento con il suo dispositivo di accesso a Internet).³²

Web semantico: applicazioni

Un gruppo di ricercatori del MIT ha sviluppato in Java

Haystack, un software gratuito che si basa sui principi del web semantico e rappresenta un esempio di interfaccia utente che fornisce all'uomo l'abilità di interagire con RDF. Sfruttando le tecnologie relative che permettono di gestire facilmente documenti, messaggi di posta elettronica, appuntamenti, contatti ecc., l'interfaccia di Hystack è in grado di visualizzare diversi tipi di informazione e non lascia trapelare il fatto che il modello di dati è rappresentato in RDF. Presentare le informazioni in modo familiare e intuitivo per l'utente è un punto chiave, poiché pochi utenti conoscono il vocabolario ontologico e la logica descrittiva: le applicazioni per l'utente finale devono essere sviluppate in modo trasparente per l'utente.

Collegando nello stesso ambiente impegni, agenda, gestione della posta elettronica e messaggistica istantanea, descrive ogni documento che risiede nel computer utilizzando RDF, riconoscendo o associando i metadati ai documenti, mettendoli in relazione in una base dati unica che rispetta RDF. Per trattare questi metadati e realizzare le relazioni, il gruppo diretto da David R. Karger, che lavora sul progetto dal 1997, ha sviluppato un proprio linguaggio di programmazione, *open source* come RDF, denominato Adenine, grazie al quale con un clic con il tasto destro del mouse sul nome del mittente di una mail si visualizzano tutte le azioni potenzialmente associate a un nome: invitare a una chat, aggiungere all'agenda, attribuire una categoria ("autore", "genitore" o "collega" ecc.). L'elenco delle azioni possibili è sempre lo stesso, indipendentemente dall'applicazione utilizzata, ed è organizzato per grandi temi (azione sulla persona, azione sul documento ecc.). I ricercatori hanno definito un'ontologia, espressa tramite l'interfaccia grafica Ozone che può essere usata per codificare la presentazione e il contenuto della pagina in modo molto più flessibile rispetto all'html: essa consente una visualizzazione molto flessibile dei dati e permette all'utente di posizionare ogni applicazione dove preferisce sul desk top di Haystack. La seconda versione del programma, ancora in fase di sviluppo, è stata messa online nell'agosto 2003: pur essendo più stabile della precedente (uscita a giugno) dimostra una certa lentezza e richiede un ambiente hardware di livello elevato.³³

James Hendler,³⁴ tra i numerosi progetti legati al Semantic

²⁹ Intervista con Derrick de Kerckhove di Raffaella Viglione, *Nel web l'individuo e la massa non si oppongono più, convivono*, "Telega" 17/18, estate/autunno 1999, <<http://www.fub.it/telema/TELEMA18/DeKerc18.html>>. Secondo de Kerckhove, la nuova ecologia delle reti si basa su tre condizioni fondamentali: l'*interattività*, il collegamento fisico della gente o delle strutture basate sulla comunicazione; l'*ipertestualità*, il collegamento delle strutture di contenuti; la *connettività* (o *webness*), il collegamento mentale della gente.

³⁰ La rete semantica può determinare i diritti dell'utente di accesso ai dati tramite l'identificazione biometrica, e successivamente personalizzare le informazioni raccolte sul web wireless (in vari formati: immagini, suoni, documenti, slide di Power Point e e-mail).

³¹ Se un utente sta cercando di acquistare qualcosa sul web, oppure cerca informazioni, l'XML può mettere assieme tutti i dati rilevanti provenienti da diverse fonti e presentare i contenuti desiderati in modo che si adattino al dispositivo mobile.

³² Esistono applicazioni di commercio mobile che usano le capacità della rete semantica per mettere insieme informazioni variabili a seconda del contesto, basate sulle preferenze personali.

³³ È richiesto almeno un Pentium III a 700 MHz con 512 Megabyte di Ram e l'installazione obbligatoria di Java 2 Standard Development Kit (che occupa da sola più di 300 Megabyte).

³⁴ Dept. of Computer Science, UM Institute for Advanced Computer Studies, Institute for Systems Research, e Dept. of Electrical Engineering University of Maryland at College Park, director of Semantic Web and agents research, Maryland Information and Network Dynamics Laboratory Acting Head, Advanced Information Technology Laboratory Head, Parallel Understanding Systems Laboratory (former Head, Autonomous Mobile Robotics Laboratory), <<http://www.cs.umd.edu/~hendler/>>.

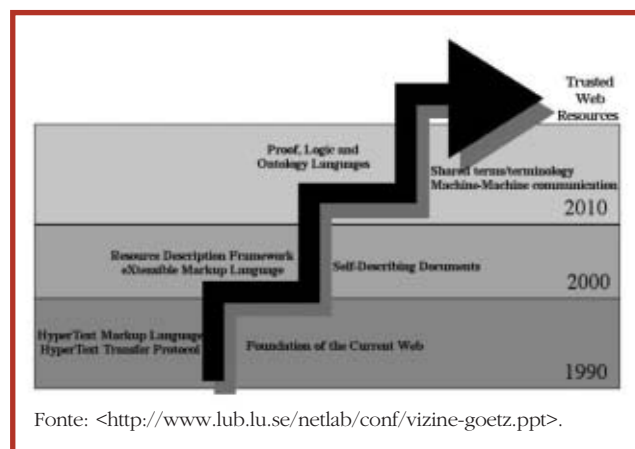
Web di cui si occupa, ha messo in rete un prototipo di navigatore semantico realizzato dall'Università di Milano e dall'Istituto superiore universitario di formazione interdisciplinare (ISUFI) di Lecce,³⁵ che pur disponendo attualmente di pochi dati è interessante come modello ontologico. In questo esempio le ontologie sono utilizzate per definire metadati che descrivono testi. Partendo da concetti semplici, si può raffinare la ricerca esprimendo vere e proprie asserzioni composte da un soggetto, un predicato e un oggetto, arrivando a formulare una richiesta del tipo: *biblioteche* (soggetto) *che hanno come servizio* (predicato) *la fornitura di documenti* (oggetto). L'approccio è completamente diverso da quello di un'interrogazione a un motore di ricerca, in cui in realtà si possono proporre tutti e tre i concetti (*biblioteca servizio documenti*) senza però poter esprimere il legame fra di essi.

Nell'ambito del progetto europeo Mesmuses, realizzato con il contributo della Commissione europea nel contesto del Quinto Programma quadro di ricerca e sviluppo tecnologico (1998-2002), nel giugno 2003 è stato organizzato il workshop "Semantic Web for scientific and cultural organisations: results of some early experiments", in cui è stato presentato il progetto che applica la metodologia del web semantico per la realizzazione di portali per musei scientifici, basandosi per la rappresentazione della conoscenza sul modello dei dati C-Web,³⁶ una delle prime implementazioni dei cosiddetti *semantic-Web services*.³⁷ Questa applicazione riesce a ottimizzare la visibilità e la recuperabilità delle risorse attraverso un'architettura funzionale basata su tre categorie di metadati (gestionali, semantici, di scopo per i diversi punti di vista) memorizzati sia localmente sia sui server web (con le risorse stesse) in uno o più depositi e scambiati tramite il protocollo WebDAV, un'estensione di http che permette agli utenti di editare e gestire file su server remoti in un ambiente collaborativo.³⁸

La comunità informatica, molto affascinata dalle potenzialità del web semantico, si chiede quanto tempo servirà per mappare tutto il web sulle ontologie e, qualora ciò accadesse, come far comunicare fra loro ontologie diverse, dal momento che c'è il forte rischio che un'ontologia sia valida solo per il dominio per il quale è stata progettata.

Riserve sull'applicabilità del Semantic Web sono state avanzate anche in occasione della tappa italiana del W3C Semantic Tour, svoltasi a Roma nel giugno 2003, soprat-

Fig. 5 – Sviluppo dei linguaggi del web nei prossimi anni



tutto dal punto di vista dell'affidabilità dei risultati. Massimo Marchiori (W3C-MIT) ha posto la fiducia al vertice del modello web presentato: ci si può fidare dell'informazione recuperata tramite queste applicazioni innovative oppure un eccesso di web semantico può trarre in inganno se se ne fa un uso non corretto?³⁹ Anche dall'intervento di Jeremy J. Carroll (HP Labs) è emerso un punto di vista piuttosto critico nei confronti del Semantic Web, di cui sottolinea la natura di puro investimento, allo stato attuale, in vista di ritorni futuri piuttosto incerti.⁴⁰

Pensare al web come a un'infrastruttura regolata da una struttura semantica, al momento significa delineare una prospettiva incerta, ma parlare dell'uso di schemi semantici all'interno di architetture legate a un dominio ben definito, quindi a un servizio per comunità ristrette di utenti orientati ad uno scopo preciso, tipo la comunità delle *digital libraries*, pone prospettive immediate più chiaramente visibili.⁴¹

La significativa immagine riportata in figura 5 sullo sviluppo dei linguaggi del web nei prossimi anni, tratta da *Semantic Web overview* di Diane Vizine-Goetz dell'OCLC Research, presenta gli strati del web semantico costruiti come nuovi linguaggi e strumenti in XML, verso un mondo di informazioni affidabili condivise tra gruppi di utenti che collaborano.

³⁵ <<http://ra.crema.unimi.it/navigator/>>.

³⁶ C-Web Functional Specification (V1.0), <[http://cweb.inria.fr/Resources/C-Web%20Functional%20Specification%20\(V1_0\).htm](http://cweb.inria.fr/Resources/C-Web%20Functional%20Specification%20(V1_0).htm)>.

³⁷ <<http://galileo.imss.firenze.it/mesmuses/indice.html>>.

³⁸ <<http://www.webdav.org/>>.

³⁹ In effetti l'informazione sul web non è sempre trasparente, ma può essere soggetta a diversi tipi di cosiddette *fratture*, quali i criteri per influenzare il ranking dei motori di ricerca, le *pop-up windows* o i *web-bugs*, ossia elementi grafici su una pagina web o un messaggio e-mail con la funzione di monitorare chi sta leggendo la pagina o il messaggio. È spesso invisibile, perché ha di solito la dimensione di 1 pixel ed è rappresentato come tag html IMG.

⁴⁰ DANIELA CANALI, *Verso il Semantic Web. "La tappa italiana del W3C Semantic Tour, "Biblioteche oggi"*, 21 (2003), 9, p. 59-62.

⁴¹ In particolare per l'influenza del Semantic Web sulle biblioteche digitali e viceversa, vedi ERIC MILLER, *Digital Libraries and the Semantic Web*, European Conference on Research and Advanced Technology for Digital Libraries, September 4-9, Darmstadt, Germany, <www.w3.org/2001/09/06-ecdl/slide1-0.html>.