

## II WEB SEMANTICO

*"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."* Così Tim Berners-Lee, inventore del Web e fondatore del World Wide Web Consortium, nel maggio 2001 definiva l'avvento del "nuovo Web", vale a dire come di una rivoluzione "semantica" per il recupero in rete d'informazioni strutturate in senso compiuto migliorando enormemente il lavoro tra persone e computer. E' l'annuncio dell'ennesima innovazione nel movimentato universo dell'information technology, e questa volta interessa particolarmente coloro che svolgendo lavori di raccolta, descrizione e organizzazione delle informazioni si collocano al centro dei flussi informativi che attraversano la Rete. Le biblioteche rispondono in pieno a queste caratteristiche e da tempo sono diffusamente presenti in Internet. Tuttavia, sebbene utilizzino sempre di più quelle tecnologie d'identificazione e descrizione delle risorse digitali presenti in sistemi distribuiti in rete chiamate "metadata", è stato comunque notato che l'impatto di Internet non ha finora prodotto al loro interno quei grandi mutamenti che si invece sono verificati in altri settori, in primo luogo in quello commerciale con l'esplosione dell'e-commerce. Ed in particolare, sembra che l'innovazione telematico-digitale non abbia inciso più di tanto su quella che è ritenuta la chiave della professionalità bibliotecaria, ossia il trattamento delle informazioni mediante catalogazione.

Ma ora, con l'approssimarsi del nuovo "Web semantico", la situazione potrebbe mutare e dischiudersi sia per le biblioteche che per archivi, centri di documentazione ecc. la possibilità di un utilizzo sistematico delle enormi risorse, per di più in continuo incremento, presenti nel Web. Per ottenere questo, è necessario però che il "Web semantico" mantenga le sue promesse a cominciare dall'assunto fondamentale, architrave della costruzione teorico-tecnologica di Tim Berners-Lee, e cioè che le informazioni diventino processabili e comprensibili direttamente alle macchine. Si tratta, con tutta evidenza, di una sorta di "rivoluzione copernicana". Infatti, fino a questo momento il trattamento dei dati ha dovuto fare i conti con una contraddizione insita nel Web: le informazioni contenute in esso sono fatte per essere lette da esseri umani e non dai computer.

L'utente che intende concettualmente i contenuti del Web, quando poi per compiere una ricerca deve interagire con la macchina e il suo software, è costretto ha tradurre mentalmente le sue domande in una serie di parole chiave, ricavandone, inevitabilmente, lunghe liste di pagine web nelle quali, è vero, compaiono i termini indicati, ma quasi sempre fuori contesto e quindi spesso inutilizzabili.

Per risolvere il problema, l'architettura del Web semantico prevede di collegare le informazioni presenti in Internet a concetti astratti organizzati in gerarchie (dette ontologie), tante quanti possono essere i principali domini applicativi d'interesse: botanica, medicina, architettura, storia, ecc. Fatto questo, le informazioni residenti in rete assumeranno un aspetto ontologico, ossia acquisteranno "valore in sé", e mediante opportune regole d'inferenza diverranno indipendenti dalle diverse forme di codifica utilizzate realizzando il sogno d'ogni "cyber-librarian": rintracciare e recuperare le informazioni disseminate nel World Wide Web esclusivamente in base al loro contenuto. Si verificherà così l'apparente paradosso di un Web che per essere più user-friendly diventerà computer-friendly, mutandosi sempre di più in un ambiente idoneo ad agenti intelligenti, instancabili programmi che lo percorreranno in lungo e largo setacciando ontologie per rispondere a domande concettualmente complesse impostate dai professionisti dell'informazione.

Se questo costituisce, a grandi linee, lo scenario prossimo venturo, tuttavia alcune tecnologie del Web semantico sono già tra noi, pressoché operative. Sono oggetto di studi, ricerche e progetti che coinvolgono, anche in forme collaborative, biblioteche e Web Consortium, e soprattutto rappresentano per gli addetti ai lavori la strada che porta all'interoperabilità, in altre parole l'agognata creazione di uno spazio informativo se non proprio "semantico" perlomeno aperto all'effettiva possibilità di scambiare automaticamente dati prescindendo dai sistemi diversi che affollano le reti.

Si tratta degli standard: XML, RDF e DC. I primi due sviluppati dallo stesso Web Consortium, il terzo nato dopo un importante convegno OCLC (On-line Computer Library Center) cioè nell'ambito delle reti bibliotecarie. L'Extensible Markup Language (XML) è un metalinguaggio flessibile capace di rendere le informazioni autodescrittive. In sostanza, con l'HTML è stato finora soltanto possibile indicare come una risorsa si sarebbe presentata, ora invece con XML si potrà definire, con etichette personalizzate e applicate a parti di testo o addirittura ad intere pagine web, cosa sono le informazioni che si stanno inviando. Il Resource Description Framework (RDF) ha invece il compito di fornire ai linguaggi o metalinguaggi, come appunto l'XML, la cornice concettuale, il framework, per consentire l'interoperabilità tra tutte quelle applicazioni che scambiano sul Web informazioni comprensibili alle macchine, insomma una piattaforma per facilitare l'elaborazione automatica dei dati all'interno della Rete Semantica. Infine, lo standard generale per i sistemi di metadata: il Dublin Core (DC), basato su un semplice impianto di soli 15 elementi offre proprio per questo il grande vantaggio di garantire l'integrazione verso ogni tipo d'accesso sia rispetto all'universo bibliotecario online degli OPAC sia riguardo alla descrizione delle risorse web nonché all'interno dei motori di ricerca aumentandone sensibilmente, se utilizzato, il livello di selezione e precisione.

Il modello operativo che, di conseguenza, sembra farsi strada è uno "stack" imperniato su questi tre standard (XML / RDF / DC), ovvero uno schema d'utilizzo paragonabile ad un'unica applicazione composta di strati diversi all'interno della quale l'XML svolgerebbe il ruolo di sintassi di base, l'RDF quello di descrivere la struttura delle relazioni e al Dublin Core spetterebbe la definizione dei contenuti.

L'interazione, intorno alla quale si sta lavorando, tra metadata e linguaggi del nuovo Web lascia prevedere importanti novità nel campo della diffusione della conoscenza e l'inizio di un'ulteriore fase di razionalizzazione per tutte le entità coinvolte. Da parte loro le biblioteche hanno cominciato a muoversi, ad interrogarsi rispetto alle nuove e complesse tematiche d'affrontare, a scambiarsi informazioni ed esperienze. Presupposti che favoriscono il nascere d'iniziative tra cui, in ordine di tempo, l'importante seminario tenutosi proprio a Roma nell'aprile scorso dall'ELAG (European Library Automation Group) dal titolo: "Semantic Web and Libraries" e che continua a proporsi come punto di riferimento mettendo a disposizione di tutti gli interessati una serie di pagine web per consultare l'intera documentazione relativa ai lavori svolti.

La visione di Tim Berners-Lee di una rete semantica popolata di macchine intelligenti è per il momento abbastanza futuristica, nondimeno gli sforzi in quella direzione presuppongono traguardi intermedi effettivamente raggiungibili e non di minore importanza: primo fra tutti quello dell'interoperabilità, chiave di volta per la biblioteca nella sua versione digitale che oggi si trova nella posizione ideale per integrare il maggior numero di risorse dilatando le proprie competenze ben oltre quei materiali "canonicamente" non bibliotecari sino ad arrivare a comprendere l'iper-massa documentale del Web.

**Fabio Di Giammarco**