



Il Web compie 20 anni. Dov'è il Semantic Web?

W3C ITALIA
Oreste Signore

CROSS KM

Semantic Web

Sommario

Risale a vent'anni fa la prima proposta del web, ad opera di Tim Berners-Lee. Oggi, che il web è un immenso contenitore di conoscenza, si parla sempre più spesso di Semantic Web, mentre il Web 2.0 costituisce uno dei fenomeni sociali più interessanti. In questo lavoro si ripercorre la storia del Web, evidenziando come già la proposta originaria contenesse i principi dell'interazione sociale e del collegamento semantico tra le informazioni. Al di là dell'enfasi posta oggi sulle reti sociali come strumenti per produrre e condividere conoscenza, vanno considerate con attenzione le potenzialità del Semantic Web e il grado di maturità raggiunto dalle varie tecnologie in quest'area, per scegliere le strategie più adeguate per innovare, salvaguardando gli investimenti.

1 Introduzione

Nel 1989 Tim Berners-Lee formulava la prima proposta del web, inizialmente ignorata, e poi accolta da un immenso successo che ha portato ad una nuova concezione dell'accesso all'informazione. Se alcuni anni fa si diceva: "*L'ho sentito in televisione*", ora sempre più spesso si dice: "*L'ho trovato sul Web*", oppure "*Lo cerco sul Web*" chiara dimostrazione di come il Web sia ormai percepito come un serbatoio di conoscenza.

La recente esplosione dell'interesse per i contenuti generati dagli utenti ha determinato un fenomeno travolgente noto con il nome di Web 2.0. In questo contesto il patrimonio informativo disponibile sulla rete cresce enormemente, e una frase ricorrente adesso è: "*Lo metto sul Web*".

Non è però facile sfruttare adeguatamente l'enorme patrimonio informativo esistente sul web, perché le informazioni sono archiviate in maniera eterogenea. Il Semantic Web si propone come l'ambiente in cui superare queste difficoltà, consentendo il reperimento, l'integrazione e l'utilizzo delle informazioni.

In questo lavoro¹ dopo un breve richiamo alle origini del web, si discuteranno gli aspetti caratteristici del Web 2.0, per poi fare il punto sullo stato di maturità delle tecnologie del Semantic Web e sulla strategia della loro adozione nell'impresa.

¹ Questo documento è disponibile all'URI: <http://www.w3c.it/papers/km13.pdf>.
La presentazione è disponibile all'URI: <http://www.w3c.it/talks/2008/km13/>

2 Le radici del Web

Il Web non è opera di un solo individuo, ma il risultato di un lungo percorso, a cui hanno contribuito molti ricercatori, favorito dalla definizione dello standard SGML e dagli sviluppi delle reti (Arpanet, Internet), innescati dalla guerra fredda e dai primi successi spaziali della Russia, partendo dalle idee pionieristiche di Vannevar Bush sull'ipertesto, e passando attraverso l'invenzione del mouse di Douglas (Doug) Englebart e le visioni affascinanti di Ted Nelson. L'invenzione del Web, quindi, non è stata un'illuminazione improvvisa, come la mitica "mela di Newton", ma un'idea nuova, quella di poter *combinare in modo libero* le idee, stimolata da una sfida posta da un ambiente di ricerca di altissimo livello (il CERN).

Ed è così che già nel 1980, in occasione del suo primo soggiorno al CERN, Tim Berners-Lee comincia ad immaginare il web (*"Suppose all the information stored on computers everywhere were linked. Suppose I could program my computer to create a space in which anything could be linked to anything."*) e realizza ENQUIRE, successivamente andato perduto, che prevedeva un insieme di *nod*i e di *link*, in cui ogni nodo aveva un *nome*, un *tipo* e una lista di *link* *tipati bidirezionali*. Nel 1989 Tim Berners-Lee propone la condivisione di documenti e definisce una DTD SGML, che segna la nascita di HTML. È del marzo 1989 *"Information Management: A Proposal"*², in cui viene delineata più compiutamente l'architettura di quello che sarebbe diventato il web. In questa proposta, inizialmente rimasta senza risposta, e ripresentata nel 1990, emergono i tre elementi fondamentali del web: universal resource identifier, protocollo per il fetch dei documenti e document mark-up (URI, HTTP e HTML).

3 Il Web 2.0

Non esiste una definizione precisa e univoca di Web 2.0. Il termine è stato coniato da Tim O'Reilly, il fondatore della nota casa editrice scientifica, che ha indicato alcune caratteristiche essenziali di una applicazione Web 2.0:

- è una applicazione in ambiente Web;
- è un servizio, non un prodotto;
- non è limitata ad un singolo prodotto software o a una particolare macchina;
- è aperta e condivisa;
- gli utenti suddivisi per gruppi e l'interazione sociale sono un componente della sua organizzazione;
- gli utenti forniscono contenuto e ne incrementano il valore.

Il Web 2.0 nasce quindi ufficialmente con la definizione di O'Reilly (anche se secondo alcuni il termine fu creato da Dale Dougherty³ e reso popolare da O'Reilly), ma il suo modello è sempre stato presente nel web fin dalla sua nascita. Non si può dimenticare che già nelle prime implementazioni Tim Berners-Lee, inventore del web, aveva realizzato un browser-editor, pensando a un'architettura in cui ogni utilizzatore potesse visualizzare informazioni e aggiungerne, sia modificando la "pagina" che introducendo nuovi link.

² <http://www.w3.org/History/1989/proposal.html>

³ http://en.wikipedia.org/wiki/Dale_Dougherty

3.1 Le caratteristiche essenziali del Web 2.0

Un elemento caratteristico del Web 2.0 è il *collaborative tagging* (spesso denominato anche *folksonomy*, *social classification*, *social indexing* o in altri modi) che designa la prassi e il metodo di creare e gestire in maniera collaborativa i tag⁴ per annotare e caratterizzare i contenuti. Il *collaborative tagging* si è largamente diffuso sul web a partire dal 2004, grazie ad applicazioni denominate “*social software*” quali il social bookmarking⁵ e l’annotazione di fotografie.

Anche se sono spesso utilizzate anche in altri contesti, le folksonomy sono internet-based. L’obiettivo del *folksonomic tagging* è rendere sempre più facile ricercare, scoprire e navigare un insieme di informazioni. Una folksonomy ben sviluppata dovrebbe essere consultabile come un vocabolario condiviso sviluppato e ben conosciuto dai suoi utenti principali. Due esempi ampiamente noti e citati di siti web che utilizzano tagging basato su folksonomy sono *Flickr* e *del.icio.us*.

L’aspetto più interessante delle folksonomy risiede forse proprio nella loro inerente inversione di tendenza (o, potremmo dire, di radicale sovvertimento): rispetto ai meccanismi tradizionali di ricerca dei contenuti forniti dai siti web, le folksonomy sembrano indicare un rifiuto dello status quo dei search engine, in favore di strumenti creati dalla comunità.

3.2 Alcune considerazioni sul Web 2.0

La fisionomia dei siti web sta cambiando: non sono più sorgenti di informazioni organizzate come isole a sé stanti, ma sono sempre più ricchi di collegamenti con altre fonti informative. Gli utenti non sono più meri fruitori passivi dell’informazione, ma diventano utenti attivi che creano e aggiungono contenuti. La differenza tra Web 1.0 e Web 2.0 è ben rappresentata dalla distanza tra la versione online della *Encyclopedia Britannica*, che costituisce un tipico esempio di Web 1.0, e Wikipedia⁶, che viene considerata ugualmente autorevole⁷, e costituisce un tipico esempio di Web 2.0.

Un altro fenomeno importante è la crescita delle reti sociali (*social networking*). Non si tratta di una novità assoluta, ma di un’evoluzione di alcune forme di interazione sociale che il web ha sempre supportato (computer conferencing, email, liste di distribuzione, etc.). La novità del Web 2.0 risiede nella diffusione di siti di social networking.

I benefici del Web 2.0 sono abbastanza ovvi: gli utenti non sono più confinati ad un ruolo meramente passivo di fruitori dei contenuti, ma partecipano in maniera attiva alla loro creazione e aggiornamento, sviluppando l’abitudine e le competenze per il lavoro di gruppo.

Tuttavia, i critici del contenuto generato dagli utenti segnalano il venir meno delle fonti tradizionali del sapere, come l’autorevolezza e il contributo degli studiosi, e segnalano come il passaggio dal mondo della carta a quello assai più dinamico, ma anche volatile, dell’informazione elettronica, determini una perdita di credibilità e fiducia nelle informazioni.

⁴ Un **tag** è una *parola chiave* (significativa) associata o assegnata ad un elemento informativo (una foto, una mappa geografica, una voce di un blog, un video clip, etc.), che descrive quell’elemento informativo e ne consente la classificazione e la ricerca.

⁵ Vedi http://en.wikipedia.org/wiki/Social_bookmarking.

⁶ <http://en.wikipedia.org/>

⁷ Non mancano tuttavia esempi di voci di Wikipedia non del tutto esatte e complete.

Un ulteriore elemento da non trascurare è il fatto che il contenuto viene creato dagli utenti utilizzando sistemi diversi (podcast, blog, wiki, sistemi di chat, e altro software per il social networking), rendendo difficile tener traccia di dove si trovi l'informazione, e problematico accedere ad essa, sia per utenti abituali che casuali. Ne consegue la necessità di disporre di strumenti che rendano più agevole all'utente la ricerca e l'integrazione dell'informazione frammentata.

Infine, vale la pena di riflettere su alcune considerazioni relative alla democrazia digitale, (Magnolfi⁸):

Nell'era del 2.0 c'è una voglia di partecipazione inedita degli utenti in rete, che può rappresentare una straordinaria opportunità di rilancio della partecipazione democratica, di allargamento di forme di democrazia deliberativa. Ma può anche scivolare verso derive populiste e verso ambigue tentazioni di democrazia diretta. Ora, se la politica e le istituzioni non definiscono regole per consentire agli utenti di prendere parola, di avere voce e nuove forme di cittadinanza attiva che possono essere molto utili al processo decisionale, questa occasione di allargamento della democrazia, se non gestita, può ritorcersi contro le istituzioni e quindi paradossalmente contro la democrazia.

3.3 Vantaggi e problemi

I vantaggi derivanti dall'adozione dell'approccio Web 2.0, sono essenzialmente legati al miglioramento della qualità dei tag e alla crescita di un senso "sociale".

Tuttavia, non possono essere trascurati anche alcuni problemi. Come evidenziato da Hardman, il Web 2.0 presenta alcuni rischi intrinseci. Proprio il fatto di contribuire in maniera così consistente al contenuto di un sito, con un evidente impegno e dispendio di energie, rende l'utente "dipendente" dal sito medesimo. L'utente è legato definitivamente al formato dati adottato, esattamente come accade nel caso in cui si utilizzino programmi proprietari. Eventuali cambi di ambiente saranno inevitabilmente onerosi. È stata proprio questa una delle motivazioni che ha portato alla nascita, ormai dieci anni fa, di eXtended Markup Language (XML). XML permetteva appunto di ridurre il rischio di rimanere legati ad un particolare formato dati, perché basato su una rappresentazione standard dei dati, che potevano poi essere utilizzati in varie maniere.

Solo a titolo di esempio, si pensi a cosa potrebbe succedere ad un utente che utilizza un particolare sito web per la condivisione delle foto, su cui carica magari migliaia di foto, tutte corredate degli opportuni tag. Cosa accade se successivamente si rende disponibile un servizio migliore, o se il sito utilizzato per la condivisione delle foto chiude, come in effetti è accaduto ad alcuni siti musicali che utilizzavano le tecnologie del Web 2.0? Con molta probabilità l'intero lavoro svolto andrà perso, o si renderà necessaria una conversione laboriosa, spesso manuale. Analogamente, non è raro il caso in cui l'utente decide di partecipare a più reti sociali, e in tal caso, in assenza di standard, potrebbe essere costretto a ripetere il lavoro varie volte.

Sono questi tipici esempi della legge di Metcalfe, secondo la quale il valore di una rete è proporzionale al quadrato del numero di nodi presenti. Ne consegue, evidentemente, che se si divide una rete in due parti uguali, il valore di ognuna di esse è un quarto dell'originale, e il valore totale la metà.

Quindi, l'esistenza di un unico World Wide Web è un fattore positivo, mentre il Web 2.0 suddivide il Web in una serie di sub-Web dedicati ad aspetti particolari, con

⁸ Sono state apportate piccole modifiche al testo, per correggere alcuni banali refusi.

il risultato di legare gli utenti a formati particolari, e diminuendo il valore della rete nel suo complesso.

4 Il Semantic Web: alcuni richiami

Il Semantic Web fornisce una infrastruttura comune che consente la condivisione e il riutilizzo dei dati tra applicazioni, imprese e comunità. È uno sforzo cooperativo guidato dal W3C, al quale partecipano molti ricercatori e partner industriali. Il Semantic Web è basato su RDF (Resource Description Framework) e talvolta viene chiamato “web of data” per sottolineare il fatto che sul web è disponibile una enorme quantità di dati, ma, dato che questi sono controllati dai programmi applicativi, e ogni applicazione li gestisce in maniera diversa, non è possibile combinare le varie informazioni, se non con un intervento manuale.

Il Semantic Web presenta due aspetti. Uno è la definizione di formati comuni per poter integrare e combinare dati provenienti da diverse fonti, mentre il Web originale focalizzava la sua attenzione sullo scambio di documenti. L'altro aspetto concerne il linguaggio per registrare come i dati sono in relazione con gli oggetti di mondo reale. Questo consente a una persona, o a una macchina, di iniziare le sue ricerche su un database, e da lì accedere alle informazioni contenute in un numero indefinito di altri database, che sono connessi per il fatto di contenere informazioni sulla stessa cosa.

Il Semantic Web si basa sull'ipotesi che le macchine possano accedere ad un *insieme strutturato di informazioni* e a un *insieme di regole di inferenza* da utilizzare per il ragionamento automatico. La sfida del Semantic Web è fornire un linguaggio per esprimere *dati* e *regole* per ragionare sui dati, che consenta l'*esportazione* sul web delle regole da qualunque sistema di rappresentazione della conoscenza. L'obiettivo principale del Semantic Web è consentire alle macchine di estrarre la conoscenza disponibile sul Web, spesso disponibile in formati eterogenei, e combinarla per poter estrarre nuova conoscenza. Questo processo è possibile solo se si riesce a rappresentare, esportare e condividere la conoscenza, mediante una rappresentazione indipendente dallo specifico ambiente operativo. Le tecnologie del W3C (RDF, RDFS, OWL, SPARQL, etc.) consentono appunto di raggiungere questo obiettivo. È importante sottolineare che *condividere i dati e la conoscenza non comporta tradurre tutte le informazioni in RDF*.

Le tecnologie del Semantic Web hanno ormai raggiunto un buon livello di maturità, e sono ampiamente utilizzate in molte applicazioni⁹.

5 Web 2.0 o Web 3.0?

Le tecnologie semantiche sono in una fase di diffusione anche nella realtà industriale. In un suo studio (maggio 2007) Gartner, nota per la qualità delle sue previsioni a lungo termine, afferma:

"Key finding: During the next ten years, Web-based technologies will improve the ability to embed semantic structures in documents, and create structured vocabularies and ontologies to define terms, concepts and relationships. This will offer extraordinary advances in the visibility and exploitation of information - especially in the ability of systems to interpret documents and infer meaning without human intervention."

⁹

<http://www.w3.org/2001/sw/sweo/public/UseCases/slides/HTML/Slides.html>

e:

"The grand vision of the semantic Web will occur in multiple evolutionary steps, and small-scale initiatives are often the best starting points."

Vi è un consenso generalizzato sul fatto che Web 2.0 e Semantic Web (o Web 3.0) sono da considerare due approcci complementari, piuttosto che alternativi. Il Web 2.0 ha un livello d'ingresso più basso (è molto facile utilizzarlo), ma anche orizzonti abbastanza limitati (in particolare, l'approccio delle folksonomy ha dei limiti intrinseci). D'altro canto, il Web 3.0 richiede investimenti iniziali più rilevanti, e quindi presenta un livello d'ingresso più alto, ma ha un potenziale nettamente superiore.

Lo studio di Gartner contiene anche altre riflessioni interessanti, consigliando di combinare tecniche Web 2.0 e Semantic Web, prevedendo una crescita graduale del Web attuale verso il Web 3.0, che presenta costi elevati, ma anche risultati rilevanti, mediante l'integrazione di tecnologie semantiche "leggere" ma di utilizzo agevole nelle tecniche del Web 2.0.

Abbiamo già evidenziato come il Web 2.0 partizioni il Web in una serie di sub-Web dedicati ad argomenti specifici, e imponga formati proprietari, inficiando così il valore della rete nel suo complesso. Al contrario, il Semantic Web offre il vantaggio di poter avere i dati distribuiti sul Web con un approccio che esclude qualsiasi centralizzazione. La semplice aggiunta di metadati ai dati consente la realizzazione di servizi di aggregazione dell'informazione che offrono gli stessi vantaggi del Web 2.0, senza incorrere nei rischi conseguenti all'adozione di formati proprietari o addirittura di perdita di dati nel caso in cui il servizio su cui si fa affidamento scompaia.

Così come il Web 2.0 è andato al di là delle aspettative pensabili al momento della definizione del primo Hypertext Transfer Protocol (HTTP), analogamente, passando da un contesto in cui i link erano creati dalle persone ad un contesto in cui i link e le informazioni sono comprensibili alle macchine, possiamo pensare che si apriranno nuovi orizzonti e nuovi fenomeni sociali, forse anche in un futuro molto prossimo.

6 Strategie per l'adozione delle tecnologie del Semantic Web

Adottare le tecnologie del Semantic Web a livello aziendale non è impresa banale. Anche se si tratta di una scommessa vincente, vanno tenuti presenti vari fattori, per evitare che le aspettative vengano deluse, e che, con una politica certamente miope, per contenere le spese e gli investimenti iniziali a fronte di benefici previsti a lungo termine, si finisca con il rinunciare ad un approccio tecnologicamente e culturalmente innovativo.

Nell'aprile 2007 Business Week online¹⁰ ha dedicato alcuni interventi all'introduzione del Semantic Web nelle imprese. Da questi si possono dedurre alcuni "consigli per l'uso" da tener presenti al momento di introdurre queste tecnologie:

- *Partire con progetti piccoli*: è bene verificare le tecnologie del Semantic Web con progetti pilota, prima di investire risorse umane e finanziarie notevoli.

¹⁰<http://www.businessweek.com/mediacenter/video/technology/5c4c05ae279f6d20a17636fdb706daea352b1955.html>

- *Verificare le referenze*: molti system integrator non hanno reali competenze sulle tecnologie del Semantic Web. È bene acquisire persone che abbiano le necessarie competenze.
- *Prepararsi a una sfida formativa*: spesso occorre del tempo perché le persone comprendano la tecnologia.
- *Trovare un alleato*: può essere difficile individuare i benefici potenziali, perciò è bene trovare qualcuno che abbia un problema che può essere risolto con il Semantic Web, e farne un partner.
- *Non agire da soli*: il Semantic Web è complesso, ed è bene avere aiuto da qualcuno
- *Non trascurare gli aspetti di privacy*: è possibile acquisire e correlare i dati dei dipendenti, ma non bisogna dimenticare di definire le linee guida per salvaguardare la privacy.
- *Non aspettarsi la perfezione*: le tecnologie permettono di reperire e correlare le informazioni in modo più rapido, ma non sono perfette. Se le informazioni di partenza sono poco affidabili, le tecnologie sono impotenti.
- *Non essere impazienti*: uno dei primi ad adottare queste tecnologie, nella NASA, afferma che i benefici potenziali possono giustificare gli investimenti in tempo, denaro e risorse, ma occorre un impegno pluriennale per nutrire speranze di successo.

7 Conclusioni

Il Web nasce in un ambiente di ricerca di altissimo livello, e già nella sua proposta originaria contiene idee non ancora completamente realizzate. Dopo la diffusione a livello planetario del Web, oggi ampio consenso e interesse riscuote il Web 2.0, il cui aspetto più rilevante è la nascita e crescita di “reti sociali”.

Anche restano ancora aperti importanti problemi squisitamente di ricerca, molte applicazioni utilizzate quotidianamente e su larga scala si basano sulle tecnologie del Semantic Web, e siamo sempre più vicini al “web of data” in cui l’utente potrà accedere senza soluzione di continuità ad un numero indefinito di risorse informative.

L’introduzione di tecnologie semantiche nelle tecniche del Web 2.0 potrà garantire grossi vantaggi, salvaguardando nel contempo gli investimenti.

L’adozione delle tecnologie semantiche nelle imprese è un processo da affrontare con la dovuta cautela, ma è da ritenere inevitabile per chi vuole restare all’avanguardia.

Nello sviluppo del Semantic Web, il W3C, in cui tecnici e ricercatori di tutto il mondo portano il loro contributo alla crescita del Web, ha giocato un ruolo guida. C’è da chiedersi se, in un mondo così fortemente competitivo e in rapida evoluzione, le imprese che restano fuori da questo processo potranno poi giocare un ruolo importante e innovativo.



Riferimenti bibliografici

Hardman Lynda and Pemberton Steven. *The Path to Web n+1*, <http://ercim-news.ercim.org/content/view/340/536/>

Madden M. and Fox S.. *Riding the Waves of 'Web 2.0*, Pew Internet Project, 5 October 2006, http://www.pewinternet.org/pdfs/PIP_Web_2.0.pdf

Magnolfi Beatrice. *Intervento del Sottosegretario di Stato alle Riforme e Innovazioni della P.A. XV Leg.* Beatrice Magnolfi al Dialogue Forum on Internet Rights (Roma, 27 Settembre 2007 http://www.quadernionline.it/igf_2008/magnolfi.html

Metcalfe, R.. *Metcalfe's law: A network becomes more valuable as it reaches more users*, Infoworld, N. 17, 2nd October 1995

O'Reilly Tim.. *What Is Web 2.0*.
<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

Poster, M. (1990). *The mode of information. Poststructuralism and social context*. Cambridge: Polity Press.

Signore Oreste. *Semantic Web: il futuro è già qui?* - 10th Knowledge Management Forum - Siena, 24-25 Novembre 2005, <http://www.w3c.it/papers/km10.pdf>

Signore Oreste. *Knowledge creation: integrazione di HTML e Semantic Web* - The 12th Knowledge Management Forum, Milano, 26-28 Novembre 2007, <http://www.w3c.it/papers/km12.pdf>

Signore Oreste. *Strutturare la conoscenza: XML, RDF, Semantic Web* - Clinical Knowledge 2003 (1st edition) - Udine, 20-21 September 2003
<http://www.w3c.it/papers/ck2003.pdf>, <http://www.w3c.it/talks/ck2003/>

Signore Oreste. *Introduzione al Semantic Web*. Web senza Barriere 2008, Roma, 7-9 maggio 2008. <http://www.w3c.it/papers/wsb08.pdf>

Terdiman, D. (2005). *Study: Wikipedia as accurate as Britannica*, CNET News. http://news.com.com/2100-1038_3-5997332.html

(Tutti i link sono stati verificati il 31 ottobre 2008)