

Blogosfera e feed RSS: una palestra per il Semantic Web?

Gino Roncaglia¹

¹ roncagl@unitus.it, <http://www.merzweb.com>, <http://roncaglia.homeip.net/merzlog>;
Università degli Studi della Tuscia, Dipartimento di Scienze Umane, Largo Università, 01100
Viterbo, Italy.

Abstract. The paper takes as its starting points the relevant role played by (i) the education of users in adding semantic metadata to the primary information published on the Web and (ii) the availability of user-friendly tools for working with semantic metadata. In both respects, weblogs may well represent the subset of the Web offering the best opportunities for the development of the Semantic Web paradigm. In order to defend this thesis, the papers (i) takes into account the history and nature of weblogs and of the main tools available to share information within the blogosphere, i.e. RSS feeds, and (ii) illustrates how weblogs and RSS feed constitute a field in which many among the ideas constituting the Semantic Web paradigm can already be applied and tested.

Sommario. L'articolo assume come punto di partenza l'importanza che hanno, per lo sviluppo del Semantic Web, la formazione degli utenti all'uso dei metadati semantici e la disponibilità di strumenti di uso semplice e intuitivo per l'associazione di tali metadati all'informazione primaria inserita in rete. Sotto entrambi questi profili il mondo weblog si propone come un sottoinsieme del Web particolarmente interessante e adatto alla sperimentazione. Per argomentare tale tesi, l'articolo prende in esame storia e natura dei weblog e del principale strumento utilizzabile per la condivisione di informazione all'interno della blogosfera, ovvero i feed RSS, illustrando alcuni fra i motivi per i quali weblog e feed RSS sembrano oggi costituire una vera e propria palestra di sperimentazione per moti fra i paradigmi alla base del Semantic Web.

1 Introduzione

Alla base del progetto Semantic Web vi è una duplice assunzione: l'idea che sia possibile aggiungere, in maniera semplice, coerente, pertinente e sufficientemente standardizzata, metadati semantici a gran parte dell'informazione primaria inserita in rete, e l'idea che questi metadati semantici siano a loro volta suscettibili di essere gestiti, analizzati e aggregati in maniera utile e funzionale attraverso l'impiego 'intelligente' di appositi agenti software.

Entrambe queste assunzioni costituiscono vere e proprie sfide per la comunità degli utenti della rete. Più specificamente, l'individuazione di adeguate griglie di metadati costituisce una sfida alle nostre capacità di modellizzazione semantica, e la realizzazione di agenti software capaci di lavorare in maniera intelligente su tali

metadati costituisce una sfida alle nostre abilità di programmazione, e in particolare di programmazione logica.

D'altro canto, la disponibilità di buone griglie di metadati, e di agenti software capaci di riconoscerli e manipolarli, non basta: occorre che gli utenti – in primo luogo gli utenti che inseriscono informazione in rete – siano abituati a *usare* i metadati, a descrivere semanticamente (eventualmente in maniera assistita, ma comunque con rigore e coerenza) l'informazione primaria da essi prodotta, a considerare la presenza di metadati semantici adeguati come una *conditio sine qua non* per l'inserimento efficace di informazione in rete.

Vi è insomma una terza sfida, che riguarda la formazione degli utenti. Una sfida che, paradossalmente, potrebbe risultare la più ardua da vincere. Infatti, la consapevolezza dell'importanza dei metadati semantici sembra ancora assai poco diffusa. La maggior parte dei sistemi di metadati disponibili, peraltro spesso tutt'altro che adeguati, è utilizzata da comunità ristrette di utenti, concentrate in maggioranza in settori professionalmente già abituati all'uso rigoroso di metainformazione descrittiva (ad esempio, nell'ambito bibliotecario e archivistico).

Certo, la progressiva diffusione di XML fornisce uno strumento per l'implementazione di sistemi di metadati che è insieme flessibile e ben integrato con le tecnologie Web lato server. Ed è prevedibile che il rapido sviluppo di software in grado di semplificare e automatizzare la realizzazione di documenti in XML, anche sulla base di schemi e DTD specifiche, continuerà con forza ancor maggiore nei prossimi anni. Ma, nuovamente, la mera disponibilità di questi strumenti non basta a garantire il successo della ciclopica opera di 'riconversione' del Web nel segno della metainformazione semantica: occorre che gli utenti siano capaci di usarli, e trovino vantaggioso farlo.

In altri termini, il Web semantico ha bisogno, per potersi sviluppare, di conquistare consensi, e può farlo solo se è capace di trovare le proprie 'killer applications': situazioni in cui un insieme sufficientemente ampio di fornitori di informazione (e non solo quelli appartenenti a un ristretto settore professionale o aziendale) sia spinto a integrare in maniera semplice e consistente metainformazione semantica nelle risorse informative prodotte, e in cui la disponibilità di tale metainformazione offra vantaggi immediati a tutti gli utenti delle risorse informative in questione.

Credo che un settore promettente, da questo punto di vista, sia quello rappresentato dai weblog e dal mondo del giornalismo e dell'informazione on-line. Nei prossimi paragrafi cercherò di illustrare sinteticamente le caratteristiche di tale settore, discuterò il possibile interesse dal punto di vista del Semantic Web di uno strumento specifico – i feed RSS – nato e cresciuto proprio in quest'ambito, e proporrò qualche considerazione generale sul rapporto fra 'blogosfera' e Semantic Web.

2 Weblog e blogosfera¹

L'esplosione del fenomeno weblog è strettamente legata all'evoluzione tecnica e stilistica di due strumenti: le pagine personali da un lato, e i programmi CMS (Con-

¹ Per un approfondimento dei temi trattati in questa sezione si rimanda al capitolo dedicato ai weblog in Calvo, M., Ciotti, F., Roncaglia, G. e Zela, M.A. (2003). Da tale capitolo (pp. 269-296) sono riprese alcune delle definizioni e delle tesi riportate di seguito.

tent Management System), nati per semplificare l'inserimento, la strutturazione e l'impaginazione dei siti Web, dall'altro.

Cominciamo dai siti personali. Molto spesso, fin dagli albori del web, essi propongono qualcosa che assomiglia alla lontana - e talvolta non troppo alla lontana - alle pagine di un diario: una serie di annotazioni, che in mancanza di un termine più adatto² chiameremo 'articoli', ciascuna delle quali può essere dedicata ad argomenti diversissimi (un avvenimento, un'osservazione, una lettura, un'esperienza...). Ogni articolo è identificato di norma (proprio come le pagine di un diario) anche o soprattutto attraverso la data nella quale è stato scritto. Molto spesso - ma, anche in questo caso, non sempre - questi articoli, essendo destinati a una sede pubblica quale è il Web, propongono segnalazioni che l'autore ritiene possano essere utili agli altri: il libro che ha appena letto è interessante (o non lo è) per questo o quel motivo, il film che ha visto o il disco che ha comprato sono o non sono piaciuti, e così via. E queste segnalazioni, a loro volta, riguarderanno spesso (ma non sempre...) risorse disponibili su Web: pagine di altri utenti, siti che sembrano particolarmente utili o interessanti, articoli o interventi reperibili in rete...

Se mettiamo per un momento da parte la questione dei contenuti, e ci concentriamo sulla struttura dell'informazione, siti di questo tipo hanno una forma abbastanza costante: articoli caratterizzati da una data di pubblicazione (e cioè la data nella quale sono scritti e inseriti in rete) ed eventualmente da un titolo, ordinati dal più recente al meno recente. Infatti, mentre la nuova pagina di un diario su carta segue quella precedente, nei 'diari in rete' - sfruttando la fluidità caratteristica del testo elettronico - tendiamo ad aggiungere il testo più recente *sopra* quello meno recente, in modo da metterlo in evidenza e far notare per primi ai nostri visitatori i contenuti che abbiamo aggiunto per ultimi e che sono dunque più aggiornati.

Spesso (ma non sempre) questi articoli conterranno anche collegamenti ad altre pagine Web.

E' proprio questo tipo di organizzazione dei contenuti che costituisce il primo elemento caratterizzante dei weblog. In un weblog tuttavia tale organizzazione non è occasionale ma si trasforma in struttura portante del sito, e si arricchisce di ulteriori e specifiche funzionalità. Innanzitutto, un meccanismo di archiviazione: dopo una certa permanenza sulla home page del weblog, gli articoli - sostituiti progressivamente dagli interventi più recenti - vengono trasferiti in pagine d'archivio. Proprio per questo, il link alle pagine d'archivio - spesso nella forma di un piccolo calendario navigabile - è presente nell'home page della quasi totalità dei weblog. Ogni articolo ha però di norma, fin dal suo primo inserimento nel weblog, anche un proprio indirizzo specifico, che non cambia col tempo ed è per questo motivo denominato *permalink*. Il *permalink* permette di far riferimento all'articolo con la sicurezza che l'utente lo potrà comunque reperire, anche quando sarà scomparso dalla home page; in questo modo, è possibile ad esempio far riferimento - all'interno di un weblog - a un articolo inserito da qualcun altro all'interno di un altro weblog. I rimandi da un weblog all'altro sono frequentissimi, e danno vita a una vera e propria ragnatela di riferimenti

² L'inglese utilizza spesso per le singole annotazioni di un diario il termine 'entry', che - in quest'accezione - non ha un corrispettivo diretto in italiano (noi parliamo talvolta di 'pagine' di un diario, ma nel contesto del Web questo termine rischierebbe di risultare fuorviante). Così come manca un corrispettivo italiano diretto per il termine 'post', spesso usato nel caso dei weblog e dei forum Web.

incrociati³. Inoltre, molti weblog dispongono della capacità di inserire e gestire eventuali commenti dei lettori ai singoli articoli. Il mondo dei weblog si trasforma così in uno spazio condiviso, popolato da utenti che dispongono di strumenti simili (e soprattutto: strumenti *compatibili*) e li utilizzano non solo per scambiarsi informazioni ma anche per approfondirle collaborativamente e per discuterle. A questo spazio condiviso è stato dato il nome suggestivo di *blogosfera*.

Se consideriamo le caratteristiche proprie dei CMS, e cioè dei sistemi di gestione dei contenuti, noteremo subito come una struttura di questo tipo sia particolarmente adatta ad essere gestita da un CMS, e da un CMS piuttosto semplice. Si tratta infatti di una struttura semplice e uniforme, che rimane costante: a variare saranno i particolari contenuti di ogni articolo, e la sua lunghezza. Non stupirà dunque che siano nati, e si stiano sviluppando con grande rapidità, sistemi di gestione dei contenuti espressamente dedicati alla realizzazione e gestione 'semplificata' di weblog. Proprio la disponibilità di tali strumenti ha contribuito all'enorme successo che il 'modello weblog' ha avuto negli ultimi mesi: sono infatti ormai attivi centinaia e centinaia di migliaia di weblog (il solo servizio offerto da Blogger⁴ ha già superato il milione di utenti).

I weblog possono essere differenziati per natura e tipologia sia delle informazioni contenute, sia degli strumenti CMS utilizzati, e io stesso ho proposto un possibile schema di classificazione⁵. A interessarci più direttamente in questa sede è tuttavia la questione dell'associazione di metadati semantici ai singoli articoli di un weblog. Per approfondire questa tematica, tuttavia, occorre presentare uno strumento specifico, che rappresenta a mio avviso una fra le novità più interessanti apparse negli ultimi anni su Web: i feed RSS.

3 Syndication e feed RSS

L'acronimo RSS può essere sciolto come 'Really Simple Syndication'⁶, e per capire di cosa si tratti può essere forse utile cominciare proprio dal concetto di *Syndication*. Dietro questo concetto c'è qualcosa di molto simile a un'agenzia di stampa. Un'agenzia produce un flusso di notizie, organizzate in base a una struttura comune: un titolo, il corpo della notizia, data e ora di emissione, e magari l'indicazione dell'autore e di una categoria tematica. Le notizie di agenzia sono riprese ed eventualmente

³ Esiste un meccanismo, denominato *trackback*, che permette di tenerne traccia: se due weblog (chiamiamoli per comodità A e B) utilizzano programmi di gestione capaci di trackback, quando un articolo sul weblog A fa riferimento a un articolo del weblog B, il programma di gestione del weblog A 'comunica' in maniera automatica al programma di gestione del weblog B che quel determinato articolo è stato citato, e il programma di gestione del weblog B aggiunge - sempre in maniera automatica - l'indirizzo dell'articolo del weblog A alla 'trackback list' dell'articolo originario. In questo modo, chi trova all'interno di un weblog un articolo particolarmente interessante o discutibile può sapere subito se quell'articolo è stato citato o discusso da altri, e dove.

⁴ <<http://www.blogger.com>>. A conferma del ruolo strategico ormai assunto dai weblog nel panorama delle risorse Web, Blogger è stato recentemente acquistato da Google.

⁵ Calvo, M., Ciotti, F., Roncaglia, G. e Zela, M.A. (2003), 275-284.

⁶ Ma anche come 'RDF Site Summary' o 'Rich Site Summary'. Torneremo in seguito su alcune fra le ragioni delle differenze nello scioglimento dell'acronimo.

pubblicate dai giornali 'abbonati' all'agenzia in questione. L'agenzia non si preoccupa di 'impaginare' la notizia: una stessa notizia potrà essere ripresa e impaginata in forme diverse da giornali diversi (mentre resteranno costanti la sua struttura e il suo contenuto). Per 'Syndication' si intende originariamente proprio la possibilità offerta da un fornitore di contenuti informativi - ai propri abbonati o a chiunque - di ricevere automaticamente ed eventualmente ripubblicare le informazioni da esso prodotte.

Quando questo meccanismo si sposta sul Web, non riguarda più solo le agenzie di stampa: ogni tipo di sito che produca informazioni e notizie può consentire - a pagamento o gratuitamente - di ricevere automaticamente ed eventualmente ripubblicare i propri contenuti informativi. Perché questo sia possibile in maniera automatica, occorre evidentemente che le notizie siano strutturate in maniera uniforme, e che le loro varie 'componenti' (titolo, corpo, data e ora, ecc.) siano opportunamente marcate.

Il formato RSS (o meglio, come vedremo, l'insieme di formati che si contendono la sigla RSS⁷) nasce proprio per marcare articoli e notizie destinati a essere scambiati via Web. Una marcatura finalizzata non all'impaginazione delle notizie (se necessaria, l'impaginazione sarà operata autonomamente dai siti 'abbonati') ma alla loro strutturazione logica. Un feed RSS dunque non è altro che un file XML contenente un flusso strutturato di notizie (ognuna delle quali costituisce un 'item'), messo a disposizione da un fornitore di contenuti informativi e pronto per essere ripreso e utilizzato da altri.

Anche se è possibile realizzare 'a mano' un feed RSS, normalmente è il sistema di gestione dei contenuti utilizzato dal sito a generarlo automaticamente, salvandolo in una apposita pagina Web (o meglio, in un file XML accessibile via Web) e aggiornandolo ogni volta che viene aggiunta una nuova notizia. Di norma, un feed RSS comprende un massimo di una quindicina di notizie: quando viene aggiunta una notizia nuova, viene anche cancellata la più vecchia fra quelle presenti nel feed.

Ma cosa c'entrano i feed RSS con i weblog, e quali strumenti ha a disposizione l'utente per 'abbonarsi' ai feed RSS dei siti che lo interessano?

La risposta alla prima domanda è semplice: i post di un weblog assomigliano molto da vicino, dal punto di vista strutturale, alle notizie di un'agenzia informativa: hanno un titolo, un contenuto, una data ed ora di pubblicazione, un autore... Il formato RSS è dunque perfetto per un weblog che voglia offrire all'esterno i propri contenuti non solo nella forma graficamente più piacevole ma assai meno 'riciclabile' di una pagina Web, ma anche in una forma più grezza: una forma che conservi la strutturazione logica ma non l'impaginazione, e permetta dunque agli utenti di raccogliere, organizzare, leggere, analizzare ed eventualmente ridistribuire le notizie apparse sul weblog in maniera totalmente indipendente dai vincoli grafici e dalle mille componenti accessorie, di interfaccia e di impaginazione, sempre presenti su una pagina Web.

Inoltre – ed è questo l'aspetto che ci interessa più direttamente in questa sede – un feed RSS permette di associare alle singole notizie anche metainformazioni semantiche: l'argomento trattato, una serie di parole chiave...

Prima di approfondire questo argomento, tuttavia, e prima di discutere le differenze esistenti al riguardo fra le diverse 'varianti' delle specifiche RSS, è opportuno dire qualcosa sui programmi aggregatori, che rappresentano il principale strumento

⁷ Per una introduzione completa ai feed RSS si veda Hammersley (2003).

disponibile per utilizzare i feed RSS. Si tratta sostanzialmente di programmi client, in grado di 'seguire' più feed RSS. Il sistema più semplice è quello di controllare a intervalli di tempo prefissati se il relativo file è stato aggiornato; ma esistono anche sistemi più evoluti, che consentono a chi offre un feed RSS di segnalare automaticamente agli aggregatori interessati la disponibilità di nuove notizie. L'utente indica all'aggregatore i feed RSS che intende seguire (la presenza di un feed RSS all'interno di un sito è spesso segnalata in home page attraverso un piccolo rettangolino arancione con la scritta 'XML' e/o l'indicazione del particolare formato RSS utilizzato), e l'aggregatore 'aggrega' i relativi messaggi, organizzandoli per sito di provenienza, per data, ed eventualmente (come vedremo) anche in base alle metainformazioni semantiche associate ai singoli post. Abbiamo detto che un aggregatore funziona di norma come un programma client; alcuni aggregatori possono tuttavia svolgere anche le funzioni di un programma server, 'servendo' all'esterno - come pagine Web o nuovamente come (meta)feed RSS - il risultato del proprio lavoro di aggregazione.

4 L'evoluzione di RSS e i suoi (troppi) 'dialetti'

Come accennato, RSS costituisce una famiglia di specifiche più che un formato unitario. La prima versione, denominata RSS 0.90, era stata sviluppata dalla Netscape Corporation all'inizio del 1999 e cioè nel periodo in cui - considerato lo sfavorevole andamento della 'guerra dei browser' - la Netscape stessa era impegnata nel tentativo di ridefinire la propria immagine proponendosi in primo luogo come portale. RSS nasce dunque con l'obiettivo di permettere la syndication di un portale informativo, e a questo fine la versione 0.90 utilizzava metadati descrittivi nel contesto delle specifiche RDF (*Resource Description Format*) sviluppate dal W3C.

Dave Winer, della UserLand (azienda pioniera nello sviluppo di sistemi software per la gestione di weblog e siti informativi), aveva creato già nel 1997 un formato XML denominato *ScriptingNews*, con finalità simili a quelle di RSS. Percependo l'interesse del linguaggio sviluppato della Netscape, ma convinto che l'aderenza alle specifiche RDF potesse complicare eccessivamente la struttura di RSS, compromettendone la diffusione, Winer sviluppò, prima in collaborazione con la Netscape (RSS 0.91) e poi autonomamente (RSS 0.92), una versione 'semplificata' di RSS, che negli anni seguenti doveva conoscere una discreta diffusione. RSS 0.92 sacrificava in parte la capacità di utilizzare metadati semantici, a favore della massima flessibilità e semplicità nella syndication delle notizie, considerata come la finalità principale del linguaggio.

Nel frattempo, però, di RSS aveva iniziato a occuparsi anche un nuovo gruppo di lavoro, il RSS-DEV Working Group, coordinato da Rael Dornfest (un collaboratore del Network O'Reilly). L'interesse di RSS-DEV era opposto a quello di Winer: rafforzare l'aderenza allo standard RDF, e dunque la capacità di utilizzare metadati semantici. Le specifiche RSS 1.0, sviluppate da RSS-DEV, sono dunque più vicine all'originale formato 0.90 della Netscape che alle sue evoluzioni successive ad opera di Winer e della UserLand.

A loro volta, Winer e la UserLand hanno continuato a sviluppare (non senza accenni polemici al lavoro del RSS-DEV) la loro concezione di RSS come meccanismo *semplice* per la syndication di contenuti: attraverso le successive versioni 0.93, 0.94 e

Blogosfera e feed RSS: una palestra per il Semantic Web?

2.0 – pur mantenendo ferma la rinuncia a RDF – l’RSS di Winer ha progressivamente allargato l’uso di metadati semantici per la descrizione di tema e contenuto delle notizie. Un gruppo di programmatori guidati da Rogers Cadenhead sta in questi mesi lavorando a un’evoluzione delle specifiche 2.0 che dovrebbe eliminarne alcune ambiguità, e sembra orientato ad adottare per le nuove specifiche il nuovo acronimo SSF (*Site Syndication Format*), mentre è stata avanzata anche la proposta di superare le divisioni e differenze fra i vari formati RSS attraverso un formato di syndication completamente nuovo, denominato *Atom*.

Come si vede, la situazione che ne risulta è piuttosto caotica, con una babele di versioni la cui numerazione non corrisponde affatto a uno sviluppo lineare; versioni che si possono suddividere in due famiglie (in base all’adozione o meno del Framework RDF) tra loro piuttosto lontane.

Dal punto di vista strettamente formale, e dal punto di vista dell’aderenza all’idea di Web semantico, non vi è dubbio che RSS 1.0, ovvero la versione sviluppata dal RSS-DEV Working Group⁸, sia la più soddisfacente. Uno dei moduli dichiarati da RSS-DEV è basato sullo spazio dei nomi (*namespace*) Dublin Core (<xmlns:dc="http://purl.org/dc/elements/1.1/">), permettendo l’adozione diretta nei feed RSS del popolare standard di metadati semantici sviluppato in ambito bibliotecario; in generale RSS 1.0 può essere esteso attraverso un meccanismo modulare basato su spazi dei nomi diversi. Va detto, tuttavia, che l’uso dei moduli Taxonomy⁹ e Dublin Core per RSS 1.0 – specificamente dedicati all’inserimento di metadati semantici – sembra per ora poco diffuso, e che in particolare i metadati Dublin Core, quando vengono utilizzati, lo sono solo in maniera assai limitata. Probabilmente si tratta di strumenti percepiti ancora come troppo complessi, che per aver successo devono essere in qualche modo ‘incorporati’, attraverso l’uso di interfacce utente semplici e intuitive, all’interno dei più diffusi strumenti di CMS utilizzati per la realizzazione di weblog.

Va ricordato peraltro che anche RSS 2.0 – che nelle specifiche di base¹⁰ oltre ai descrittori di base come <title> e <description> di ogni notizia prevede il solo elemento <category> per il riferimento a una tassonomia semantica – ammette nella versione 2.0 l’espandibilità attraverso il meccanismo dei moduli basati su spazi dei nomi (e dunque, ad esempio, l’uso di metadati Dublin Core). Come vedremo fra breve, esistono già proposte di moduli per RSS 2.0 specificamente orientati all’aggiunta di metainformazione semantica.

Considerando la pluralità di versioni e ‘dialetti’ che abbiamo ricordato, e tenendo presenti le polemiche che hanno accompagnato la divisione fra le due principali ‘scuole’ RSS¹¹, la semplice sopravvivenza di una qualche attenzione verso RSS sarebbe già sorprendente: in effetti, quella che abbiamo descritto sembrerebbe la storia del suicidio di un formato, più che della sua affermazione. Eppure, nonostante le polemiche e le divisioni, l’attenzione verso RSS è continuata a crescere, ed è cresciuto rapidamente ed esponenzialmente il numero dei siti che ne fanno uso. Un dato di

⁸ <www.purl.org/rss/1.0/>.

⁹ <http://web.resource.org/rss/1.0/modules/taxonomy/>.

¹⁰ <http://blogs.law.harvard.edu/tech/rss>.

¹¹ Alla luce delle considerazioni fin qui svolte, non sorprenderà che RSS 1.0 sciolga l’acronimo RSS come ‘RDF Site Summary’, mentre RSS 2.0 lo sciolga come ‘Really Simple Syndication’.

fatto che può sorprendere, ma che ha una spiegazione: per sua stessa natura, la blogosfera ha assoluto bisogno *sia* di strumenti di syndication, *sia* di strumenti per l'associazione di metadati semantici ai singoli messaggi. E ha bisogno che questi strumenti siano integrati fra loro e semplici da usare. Ecco dunque che il variegato e multiforme mondo dei weblog si è trasformato in un'immensa palestra di sperimentazione dei vari formati RSS: *Blogger* e *Movable Type* – due fra le piattaforme più diffuse per la realizzazione di weblog – adottano RSS 1.0¹², mentre le piattaforme UserLand (*Radio* e *Manila*) adottano RSS 2.0.

E' auspicabile che in futuro la frammentazione del mondo RSS possa essere superata: dal punto di vista degli utenti, infatti, la disputa sull'adozione o meno del Framework RDF può risultare un falso problema. Disponendo degli strumenti giusti, gli utenti si possono concentrare sul contenuto e sulla categorizzazione semantica dei loro messaggi, piuttosto che sulla sintassi del relativo feed RSS.

5 Blogosfera e semantic Web

Come abbiamo osservato, la blogosfera rappresenta un ambiente ideale per la sperimentazione di strumenti legati all'idea di Web semantico. Si tratta infatti di un sottoinsieme del Web caratterizzato da

1. una relativa uniformità strutturale dei contenuti;
2. una relativa semplicità strutturale dei contenuti;
3. l'uso diffuso di un gruppo abbastanza ristretto di strumenti avanzati per la creazione e gestione dei contenuti (CMS), capaci di forte interoperabilità e che nella grande maggioranza di casi fanno nativamente ampio uso di XML. Anche per questo motivo, i CMS utilizzati nella blogosfera risultano particolarmente adatti a semplificare gli aspetti 'tecnici' del processo di aggiunta di metadati semantici all'informazione primaria;
4. una buona diffusione di RSS, una famiglia di formati XML orientata alla strutturazione e condivisione dei contenuti che, nonostante le differenze fra le diverse versioni, permette comunque l'aggiunta di metainformazioni semantiche più o meno sviluppate, in maniera abbastanza semplice per l'utente;
5. la presenza di una comunità di utenti molto attiva, fortemente interconnessa, interessata allo scambio di contenuti, e assai più consapevole dell'importanza degli strumenti di gestione e classificazione semantica dell'informazione di quanto non avvenga per la maggior parte dei responsabili di altri tipi di siti Web;
6. una popolazione di utenti e un insieme di strumenti relativamente giovani e in rapidissima evoluzione, meno gravati dal consolidarsi di alcune fra le 'cattive pratiche' di gestione dei contenuti più ampiamente diffuse nel mondo Web.

In forma diversa, alcune fra queste caratteristiche della blogosfera sono richiamate da Steve Cayzer e Paul Shabajee nell'illustrare le motivazioni del loro lavoro sul paradigma del 'semantic blogging' nel campo specifico della gestione bibliografica:

It is worth examining in more detail why we are bringing two successful, but distinct, paradigms (blogging and the semantic web) together. We believe

¹² In realtà *Movable Type*, pur adottando RSS 1.0 nel template standard, consente di utilizzare anche RSS 2.0.

Blogosfera e feed RSS: una palestra per il Semantic Web?

that there are compelling reasons to combine the two. The rich structure and query properties enabled by the semantic web greatly extends the range of blogging behaviours, and allows the power of the metaphor to be applied in hitherto unexplored domains. Small group bibliographic management is a concrete example of a task that illustrates the benefit from the combined paradigm. Here is a task which is characterised by a need to share small items of information with a peer group in a timely, lightweight manner. This information should be easily publishable, easily discoverable and easily navigable. It should be simple to enrich the information with annotation, either at the point of delivery or later. The information should be archived in a commonly understood way for effective post-hoc retrieval. It should be possible to be notified, in a timely way, of new items of interest. We believe that a combination of blogging and semantic web technologies offers an ideal solution to this problem - blogging for low barrier publishing, a simple shared conceptual model, and a mechanism for natural, dynamic community formation - semantic web for rich structure, which enables richer community annotation, and rich query, which enables more powerful discovery and navigation. (Cayzer e Shabajee, 2003)

Non stupisce, dunque, che l'interesse verso il 'semantic blogging' cominci a produrre i primi frutti. Esistono programmi, come *Live Topics* (<<http://www.novissio.com/products/liveTopics/liveTopics.html>>), in grado di semplificare l'associazione di metadati semantici ai post di un weblog. Esiste una proposta di specifiche per un modulo RSS 2.0 (ENT: *Easy News Topics*) che consente in maniera assai semplice l'inclusione di metainformazioni semantiche in un feed RSS 2.0 (<<http://www.purl.org/NET/ENT/1.0/>>). Esistono strumenti lato server, come *k-collector* della Evecors (<<http://www.evecors.com/>>), in grado di utilizzare metadati di questo tipo per aggregare semanticamente in categorie i post di un weblog. Esistono strumenti che usano il meccanismo dei TrackBack per generare aggregazioni semantiche dei post di weblog diversi (*The Internet Topic Exchange*: <<http://topicexchange.com/>>). Esistono strumenti Perl in grado di semplificare la generazione di feed RSS 1.0 che utilizzano sia il modulo Taxonomy sia il modulo Dublin Core (<<http://www.webreference.com/perl/tutorial/22/>>). E' stata proposta una *Weblog MetaData Initiative* (<http://www.wmdi.org/>). Esiste un'iniziativa specifica di SWAD-Europe (*Semantic Web Advanced Development for Europe*) dedicata al semantic blogging (<http://www.w3.org/2001/sw/Europe/reports/open_demonstrators/hp-requirements-specification.html>). E così via¹³.

Nonostante l'indubbia difficoltà connessa all'individuazione di griglie efficaci e condivise di metadati semantici (è possibile che in questo caso i primi passi possano venire dai weblog aziendali, caratterizzati da una maggiore standardizzazione nella tipologia dei post), il mondo dei weblog sembra insomma costituire davvero una palestra ideale per il Semantic Web, anche perché popolato dalle persone maggior-

¹³ Per alcune altre applicazioni interessanti si vedano le pagine *Blogstreet's RSS Ecosystem* (<<http://www.blogstreet.com/rssecosystem.html>>); *Lockergnome's RSS Resource* (<<http://rss.lockergnome.com/>>); *O'Reilly's RSS Devcenter* (<<http://www.oreillynet.com/rss/>>).

Gino Roncaglia

mente interessate a questo sviluppo, e capaci di lavorarci sopra con profitto. Non resta che augurarsi che questo lavoro prosegua, e possa dare i suoi frutti¹⁴.

Riferimenti bibliografici

- Atom specifications*: < <http://intertwingly.net/wiki/pie/>> (ultima consultazione, ottobre 2003).
- Calvo, M., Ciotti, F., Roncaglia, G., Zela, M.A. (2003), *Internet 2004*, Roma-Bari, Laterza.
- Cayzer, S. e Shabajee, P. (2003), *Semantic Blogging and Bibliography Management*, <http://www-uk.hpl.hp.com/people/steve_cayzer/papers/030524blogpaper_final.pdf>, giugno 2003 (ultima consultazione, ottobre 2003).
- Hammersley, B. (2003), *Content Syndication with RSS*, Sebastopol (CA), O'Reilly.
- Hammond, T. (2003), *Why Chose RSS 1.0?*, <<http://www.xml.com/pub/a/2003/07/23/rssone.html>>, 23 luglio 2003 (ultima consultazione, ottobre 2003).
- King, A. (2001), *The Evolution of RSS*, <<http://www.webreference.com/authoring/languages/xml/rss/1/>>, 2001 (ultima consultazione, ottobre 2003).
- Nottingham, M. (2003), *RSS Tutorial for Content Publishers and Webmasters*, <<http://www.mnot.net/rss/tutorial/>>, versione 0.78, ottobre 2003 (ultima consultazione, ottobre 2003).
- Pilgrim, M. (2002), *What is RSS*, <<http://www.xml.com/pub/a/2002/12/18/dive-into-xml.html>>, 18 dicembre 2002 (ultima consultazione, ottobre 2003).
- RSS 0.90 specifications*: <<http://www.purplepages.ie/RSS/netscape/rss0.90.html>>, marzo 1999, (ultima consultazione, ottobre 2003).
- RSS 0.91 specifications*: Netscape, <http://www.scripting.com/netscapeDocs/RSS_00_91_Spec_revision_03.html>, luglio 1999; UserLand, <<http://backend.userland.com/rss091/>>, 4 giugno 2000 (ultima consultazione, ottobre 2003).
- RSS 0.92 specifications*: <<http://backend.userland.com/rss092/>>, 25 dicembre 2000 (ultima consultazione, ottobre 2003).
- RSS 0.93 specifications*: <<http://backend.userland.com/rss093/>>, 20 aprile 2001 (ultima consultazione, ottobre 2003).
- RSS 1.0 specifications*: < <http://purl.org/rss/1.0/spec>>, 9 dicembre 2000 e successive integrazioni (ultima consultazione, ottobre 2003).
- RSS 2.0 specifications*: < <http://backend.userland.com/rss/>>, 18 settembre 2002 e successive integrazioni (ultima consultazione, ottobre 2003).

¹⁴ Sono grato a Paolo Valdemarin per i preziosi suggerimenti su una prima versione di questo articolo.