

Il Semantic Web

Paolo Bouquet¹ e Roberta Ferrario²

¹ bouquet@dit.unitn.it <http://www.dit.unitn.it/~bouquet>
Università degli Studi di Trento, DIT, via Sommarive 10, 38050 Trento, Italy
² roberta.ferrario@loa-cnr.it <http://www.cs.unitn.it/~ferrix>
Laboratorio di Ontologia Applicata, CNR, via Solteri 38, 38100 Trento, Italy

Introduzione

In un articolo molto noto, pubblicato nel maggio del 2001 sulla prestigiosa rivista *Scientific American*, Tim Berners-Lee (accreditato come uno degli “inventori” dell’attuale World Wide Web), James Hendler e Ora Lassila scrivono:

“Il Semantic Web è un’estensione dell’attuale Web, nella quale all’informazione viene dato un significato ben definito, permettendo così ai computer e alle persone di lavorare meglio in cooperazione.”³

[Tim Berners-Lee, James Hendler, Ora Lassila, “The Semantic Web”,
Scientific American, Maggio 2001]

Per quanto il passo riportato proponga una sfida interessante, essa pare riguardare un ambito prettamente tecnologico, in cui ricercatori di discipline non informatiche sembrano avere esclusivamente un ruolo di utenti, o al massimo di contorno. Perciò, come introduzione a questo numero monografico di *Networks* dedicato al tema del Semantic Web, diventa essenziale tentare di rispondere alla seguente domanda: perché mai il progetto del Semantic Web dovrebbe essere di interesse per i lettori di *Networks*? Esistono ambiti di ricerca rilevanti per loro? È possibile che essi diano un contributo al progetto?

1 Dal World Wide Web al Semantic Web

Come dice la citazione riportata poco sopra, il Semantic Web non è una nuova entità indipendente o alternativa rispetto al Web come lo conosciamo oggi, ma è piuttosto una sua estensione che ne dovrebbe aumentare enormemente le potenzialità. Eppure il Web attuale sembra essere uno strumento già estremamente potente. In che senso dunque se ne vogliono accrescere le potenzialità?

Un modo semplice per rispondere a questa domanda è quello di riflettere su alcune limitazioni del Web con cui la maggior parte dei suoi utenti si scontra molto spesso. Eccone alcuni esempi:

³ L’articolo è disponibile online nella versione originale inglese all’indirizzo <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>, la traduzione della citazione in italiano è nostra.

Ricerca di documenti. Tutti ormai usiamo il Web per trovare documenti utili alla nostra vita professionale o personale. Per far ciò, sono disponibili due strade: seguire i link da una pagina all'altra fino a trovare quello che cercavamo, o servirci del nostro motore di ricerca preferito (Google, Yahoo, Altavista, Virgilio, AskJeeves, ...) per ottenere una lista di link tra i quali scegliere quelli di nostro interesse. Ognuno dei due metodi ha vantaggi e svantaggi:

- seguire i link ipertestuali è un processo cognitivamente molto ricco per noi umani, che in genere capiamo il contenuto del documento a cui punta il link dalla descrizione che ne troviamo nella pagina da cui partiamo, e da altra informazione di contesto (per esempio, se troviamo un link descritto come “Carlo Penco” nella pagina del Dipartimento di Filosofia dell’Università di Genova, sezione “People”, capiamo non solo che cliccando sul link troveremo la pagina di un tale che si chiama Carlo Penco, ma anche che egli con ogni probabilità è un membro del Dipartimento di Filosofia di Genova e non un calciatore del Trento). Tuttavia, tale processo può essere estremamente dispendioso in termini di tempo, e difficile da avviare nei casi in cui non sappiamo da dove partire per iniziare a seguire link che ci possano condurre a ciò che cerchiamo (per esempio, da dove partiamo a seguire link se cerchiamo informazione su un certo Carlo Penco, del quale però non sappiamo nient’altro se non che raccontava barzellette interessanti sulla spiaggia?);
- usare motori di ricerca ha l’evidente vantaggio di richiedere pochissima informazione in partenza (dal nome “Carlo Penco” possiamo trovare molte pagine che contengono tale sequenza di caratteri). Tuttavia, anche questo metodo ha dei limiti noti a tutti. Innanzitutto, i motori di ricerca non coprono tutto il contenuto del Web (si parla di un 80% di “hidden web”, cioè contenuto non indicizzato o persino non indicizzabile per ragioni tecniche). Ma soprattutto la ricerca per parole chiave può essere frustrante per l’elevato numero di “falsi positivi” (pagine che contengono la parola chiave da noi inserita, ma non parlano di quello che cercavamo, per esempio per problemi di polisemia) e di “falsi negativi” (nei risultati della ricerca non compaiono pagine che sarebbero state di nostro interesse, ma che non contenevano esattamente la parola chiave da noi immessa – magari contenevano un suo sinonimo)⁴.

Ricerca di informazione. Un altro forte limite del Web attuale è che la ricerca, comunque sia effettuata, restituisce sempre documenti, e non esattamente l’informazione che stiamo cercando. Per cui, se vogliamo sapere quali articoli ha pubblicato Carlo Penco nel 2003 o quelli in cui ha avuto dei co-autori, i due metodi di ricerca descritti sopra ci permetteranno di trovare la pagina delle pubblicazioni di Carlo Penco (se siamo fortunati), ma dopo ci sarà richiesto un certo lavoro manuale per creare la lista che ci interessa. Al momento attuale, nulla ci supporta in richieste come quelle descritte. Per esempio, non c’è modo di fare riferimento al concetto di “co-autore”, o di “anno di pubblicazione” di un articolo. Tali ricerche possono

⁴ Nota frustrante di uno dei curatori di questo numero: se cercate “Bouquet” con Google, le prime pagine di link – quasi tutti ormai sponsorizzati – contengono esclusivamente informazioni su fi orai in tutto il mondo. La mia pagina personale compare solo molto più in basso

solo essere fatte a mano o approssimate combinando più parole chiave, ma questo ci espone a tutti i problemi descritti sopra.

Integrazione di informazione. Spesso diverse parti dell'informazione che ci interessa sono contenute in diversi documenti, magari disponibili su siti diversi. Per esempio, se volessimo sapere quali autobus sono disponibili dall'aeroporto di Genova verso il centro città in coincidenza con un certo volo, dovremmo consultare separatamente il sito della compagnia aerea per sapere gli orari dei voli, poi quello della compagnia di trasporto terrestre per gli orari dei bus, e infine combinare manualmente le due cose in un piano di viaggio. Ancora una volta, nulla ci permette di descrivere concetti come “prima”, “dopo”, “coincidenza”, “percorso” e così via, né di combinare informazione proveniente da fonti diverse in un piano che risolva il nostro problema.

Cooperazione. Infine, è chiaro che il Web attuale non permette la cooperazione tra programmi e tra programmi e utenti umani per risolvere problemi complessi. La maggior parte dei siti web non sono progettati per fornire servizi ad altri servizi, ma semplicemente come contenitori di informazioni che possono essere estratte a richiesta. Per contro, molte applicazioni (per esempio, l'organizzazione di un viaggio) richiederebbero da un lato che i siti Web di varie organizzazioni potessero interagire in modo flessibile e dinamico (per esempio, per comporre un piano di viaggio e alloggio e un programma di visite turistiche integrate con il piano di viaggio), e dall'altro che gli utenti potessero intervenire nel processo interagendo con i programmi dei vari siti (per esempio, verificare interattivamente che il risultato sia conforme alle esigenze del potenziale viaggiatore, permettendogli di richiedere modifiche o proposte alternative nel corso del processo di definizione).

A questo tipo di problemi intende rispondere il Semantic Web. Vediamo come.

2 Il programma di ricerca

Il Semantic Web è un progetto di ricerca basato su un programma molto preciso: quello di ridefinire e ristrutturare i dati su Web in modo che il loro significato sia accessibile non solo a utenti umani (come avviene per il Web attuale), ma anche – e soprattutto – a programmi che li utilizzano per manipolarli, integrarli, renderli disponibili per diverse applicazioni – non solo mostrarli, come succede ora. Tale programma si basa su quattro pilastri fondamentali:

2.1 Rappresentazione della conoscenza

Per poter dare un risultato veramente soddisfacente per l'utente, i programmi scritti per il Semantic Web devono avere un qualche tipo di accesso al significato dei dati con cui lavorano. Per esempio, a differenza degli attuali motori di ricerca, devono poter capire se “Carlo Penco” è una persona o una marca di profumo, e cosa deriva dalla decisione tra uno dei due significati. Pertanto, l'aggettivo “semantico” nella locuzione “Semantic Web” traduce questa volontà di creare qualcosa che sia significativo per dei programmi di computer, e non solo per l'utente umano. Il primo passo per mettere le macchine in

grado di operare in maniera più efficiente sono dei documenti ben strutturati corredati da *metadati*. I metadati sono, letteralmente, “dati che parlano dei dati”, ed esprimono parte di quello che c’è da sapere su un certo insieme di dati, e ciò che da tale conoscenza è possibile inferire mediante regole di ragionamento. Questo significa che il Semantic Web intende recuperare tutta la tradizione della ricerca in Intelligenza Artificiale che va sotto il nome di *rappresentazione della conoscenza*, focalizzandola tuttavia su un nuovo obiettivo molto ben delimitato e preciso. Lo strumento tecnico elaborato per esprimere i metadati nel Semantic Web sono dei linguaggi di annotazione (o *markup language*) costruiti a partire da XML (*EXtensible Markup Language*), come per esempio il noto RDF (*Resource Description Framework*). Le annotazioni servono a rappresentare il significato dei dati annotati. I linguaggi di annotazione ci permettono quindi di annotare la stringa “Carlo Penco” con l’informazione che Carlo Penco lavora presso il Dipartimento di Filosofia dell’Università di Genova, che collabora ai progetti X, Y e Z e che insegna il corso di Filosofia del Linguaggio.

2.2 Utilizzo di ontologie

Tuttavia, l’introduzione di linguaggi di annotazione è solo il primo passo verso l’ingresso della semantica nel Web, poiché ai linguaggi da soli manca ancora un sostrato concettuale che leghi i termini tra loro, mostrandone le relazioni e facendone quindi emergere il significato. Questo compito è svolto dalle cosiddette *ontologie*, cioè documenti o file che hanno lo scopo di esprimere il significato inteso di un certo insieme di termini definendone le relazioni reciproche. Per dare un’intuizione di che cosa un’ontologia aggiunga a un linguaggio di annotazione, si può pensare alla differenza esistente tra un semplice elenco di parole italiane e un dizionario concettuale. Il primo ci dice solo quali parole è lecito usare, il secondo fornisce anche un insieme di relazioni logico-grammaticali con altri termini (per esempio, che ci dica che Carlo è un professore, che professore è un ruolo accademico e sociale, che ogni professore afferisce a un Dipartimento accademico, che i Dipartimenti sono strutture organizzative delle Università, che l’Università è un tipo di istituzione, ecc.). Le ontologie pensate per il Web hanno spesso la forma di tassonomie corredate di un insieme di regole di inferenza, e forniscono definizioni formali delle relazioni tra termini.

2.3 Agenti software

Gli agenti software sono programmi che dovrebbero sfruttare la conoscenza contenuta nei metadati per fornire servizi complessi agli utenti del Semantic Web. Per far ciò, gli agenti devono essere dotati di un buon livello di autonomia (non possono rivolgersi continuamente all’utente per decidere cosa fare) e di capacità di ragionare in modo complesso sull’informazione a cui hanno accesso. Questo significa al minimo essere in grado di rappresentare gli obiettivi di un certo utente, di mettere in atto una sequenza di azioni che possa soddisfare tali obiettivi, ed eventualmente di cooperare con altri agenti per ottenere tale risultato.

Naturalmente, ognuno dei punti elencati comporta l’emergere di molti problemi, alcuni dei quali di difficile soluzione. Il più generale di tutti ha a che fare con un concetto solo apparentemente tecnico, vale a dire in concetto di *interoperabilità*. Interoperabilità

significa non solo l'abilità di programmi progettati indipendentemente e implementati su piattaforme diverse di scambiarsi dati (cosa che già oggi in parte avviene), ma soprattutto di cooperare tra di loro sulla base di una "comprensione" del significato dei dati oggetto di scambio (in questo senso si parla di *interoperabilità semantica*). In altre parole, ciò richiede che i programmi per il Semantic Web (e in particolare gli agenti software) debbano "capirsi" senza presupporre un accordo a priori su cosa significhi un certo dato (detto altrimenti, due agenti diversi potrebbero non condividere la rappresentazione del mondo in base a cui agiscono), e su quale uso verrà fatto di un certo dato. Inoltre, agenti molto evoluti dovrebbero anche essere capaci di dedurre – a partire dall'analisi di un documento e grazie a processi inferenziali – nuovi metadati non presenti in esso, di acquisire nuove "conoscenze e capacità" quando vengono in contatto con nuove ontologie e di ricavare informazioni importanti anche da servizi che rispondano solo parzialmente alle richieste dell'agente stesso. In aggiunta, ci si aspetta che la cooperazione tra questi agenti avvenga anche in base a una capacità di ragionare sulle intenzioni degli utenti umani e degli altri agenti software, e nel rispetto di norme di tipo etico e legale.

Di fronte al panorama che abbiamo mostrato, quali sono le linee di ricerca che risultano più interessanti da un punto di vista filosofico e nelle quali un ricercatore con un *background* in filosofia potrebbe dare un importante contributo?

3 Problemi di ricerca per filosofi

A nostro parere, il Semantic Web potrebbe diventare una sorta di "laboratorio" in cui molte teorie filosofiche potrebbero essere messe alla prova o applicate, e in cui molti nuovi problemi filosofici potrebbero emergere. Di seguito, discutiamo alcuni degli ambiti che a noi paiono più interessanti e promettenti.

Innanzitutto, vorremmo citare la definizione e lo sviluppo di ontologie. Il concetto di ontologia usato in questo contesto sembra molto differente da quello filosofico, eppure il problema di fondo appare riconoscibile: quali "entità" un agente deve presupporre che esistano per poter interagire e cooperare con altre entità in un certo contesto? Quali relazioni esistono tra queste entità? Si osservi poi che esiste una distinzione tra le cosiddette ontologie "top level" e quelle di dominio. Mentre le seconde possono essere paragonate a teorie che un agente ha su un certo ambito di applicazione (per esempio, la finanza o l'informatica), le prime si propongono di definire i concetti più generali che un agente dovrebbe utilizzare, quali il concetto di "oggetto", "spazio", "tempo", "evento", "azione", "legge", e così via. È facile osservare che le ontologie *top level* manipolano concetti che sono stati ampiamente e dettagliatamente analizzati nella letteratura filosofica e sicuramente una certa dimestichezza con i relativi dibattiti filosofici può essere d'aiuto nell'affrontare questioni critiche che possono emergere in ambito ontologico. Soprattutto, ci pare che la filosofia possa dare un contributo decisivo nella scelta (e nell'analisi) delle cosiddette primitive ontologiche, nell'individuazione degli oggetti di un certo dominio e nella rappresentazione adeguata dei concetti più generali, a partire da quelli appunto di spazio, tempo, soggetto e oggetto, solo per fare alcuni esempi.

Un secondo ambito di collaborazione è quello della comunicazione (e quindi collaborazione) tra agenti con diverse "teorie del mondo". Infatti, il Semantic Web offre un

esempio concreto di come diverse comunità umane (per esempio, diversi gruppi di ricercatori) sviluppino diverse rappresentazioni persino dello stesso dominio (è esperienza comune di chi ha partecipato a qualche comitato di standardizzazione di ontologie quella di assistere all'impossibilità di mettere d'accordo persone diverse sulle primitive ontologiche da cui partire per costruire l'ontologia di un certo dominio, o sulle relazioni tra tali concetti primitivi). Se le cose stanno così, allora esiste un serio problema di far comunicare tra di loro agenti che utilizzano termini definiti su ontologie tra loro eterogenee (questo problema è spesso chiamato il problema dell'*eterogeneità semantica*). La soluzione di questo problema richiede un armamentario concettuale di cui i filosofi, soprattutto quelli che studiano il linguaggio e la comunicazione umana, sono forniti. Si tratta infatti di definire "protocolli di comunicazione" volti a trovare un accordo sul significato dei termini (talvolta chiamato problema del *coordinamento e della negoziazione semantica*), sulla traduzione da un linguaggio a un altro, di dare una definizione di livello di comprensione accettabile. Tali protocolli non possono prescindere dalla nozione di atto linguistico, e di interpretazione di un atto linguistico, includendo quindi non solo aspetti di semantica, ma anche di pragmatica filosofica.

In terzo luogo, l'accento posto nella definizione di Semantic Web sul tema della cooperazione tra agenti umani e software pone interessanti sfide rispetto ad altri temi di rilievo per la filosofia, come quelli del rapporto tra conoscenza/credenza e azione, nonché quello della definizione di un modello di agente in cui gli stati interni dell'agente – come appunto le sue credenze, i suoi desideri, le sue intenzioni – determinano (almeno in parte) il suo comportamento. Si apre quindi un'intera area di problemi in cui i filosofi possono sottoporre a verifica concreta le teorie sviluppate in quest'area, e trarre nuova ispirazione (in senso quasi galileano) da ciò che gli "ingegneri" fanno in questo campo dal punto di vista tecnologico.

Da un punto di vista più strettamente logico-formale, alcuni tipi di ragionamento che sono costantemente oggetto di analisi e sviluppo della logica matematica e filosofica e che possono essere di estrema utilità nell'implementazione di sistemi di ragionamento automatico per il Semantic Web sono le logiche non monotone, le logiche intensionali (come quelle della credenza) e in generale le logiche del senso comune.

Non dimentichiamo, infine, che il Semantic Web si propone come terreno di collaborazione fortemente multi-disciplinare. Infatti, non solo esso pone sfide interessanti per un largo numero di filosofi, ma esso richiede la collaborazione di altre discipline quali la sociologia (per esempio, per lo studio delle dinamiche sociali tra agenti software e tra essi e agenti umani, come per esempio quelle legate alla fiducia o alla delega), l'economia (per esempio, per la definizione di modelli di razionalità per agenti, nonché per lo studio dei moventi economici che determinano il comportamento di agenti autonomi), il diritto (per esempio, per problemi legali legati alla delega di decisioni ad agenti artificiali, o per problemi legati ad aspetti contrattuali e di privacy dei dati). Questo ci pare un ulteriore movente per suscitare l'interesse dei filosofi.

In conclusione, come osservano Berners-Lee, Hendler e Lassila al termine del loro articolo citato all'inizio di questa introduzione, un Semantic Web completamente realizzato potrebbe diventare uno strumento per supportare l'evoluzione dell'intera conoscenza umana. Infatti, secondo i tre autori, lo sviluppo di tale conoscenza è in una tensione costante tra la capacità di piccoli gruppi/comunità di produrre nuova conoscen-

za e di innovare in modo rapido ed efficace, e l'esigenza di far uscire queste conoscenze locali da questi piccoli circoli per renderle comprensibili e disponibili a una cerchia sempre più vasta di esseri umani. Il Semantic Web, con la sua capacità di annodare concetti creati da una certa comunità in una rete universale e accessibile a tutti si propone come una nuova famiglia di strumenti in cui gli esseri umani possono vivere, lavorare e apprendere insieme. Un progetto ambizioso, davanti al quale riteniamo che i filosofi non dovrebbero voltare la testa o tirarsi indietro, bensì giocare un ruolo da protagonisti.

4 Gli articoli di questo numero

Gli articoli raccolti in questo numero di Networks costituiscono dei significativi esempi di ricerca nei diversi ambiti del Semantic Web. Per ovvie ragioni non sono stati inclusi articoli che presentassero aspetti troppo legati alla tecnologia del Semantic Web, mentre si è cercato di dare spazio a lavori che mettessero in luce problemi potenzialmente interessanti per ricercatori di aree disciplinari non strettamente legate alla tecnologia.

Frank van Harmelen, uno dei ricercatori più impegnati a livello internazionale nello sviluppo del Semantic Web (per esempio, sarà il General Chair della prossima International Semantic Web Conference – ISWC-04 – che si terrà in Giappone), propone in forma di intervista la sua visione del Semantic Web, e dei vantaggi che esso potrebbe portare nel prossimo futuro.

Oltremari e i suoi coautori illustrano invece il loro lavoro relativo allo sviluppo di ontologie per il Semantic Web. In particolare, il contributo illustra la distinzione tra cosiddette *lightweight ontologies* (ontologie “leggere”), cioè mirate a specifici domini applicativi e *heavyweight ontologies* (ontologie “pesanti”), vale a dire ontologie fondazionali il cui scopo non è tanto quello di definire i concetti impiegati in un certo dominio, quanto quello di definire le categorie più primitive alle quali possono poi essere ricondotti tutti i concetti specifici delle ontologie leggere. Tra i problemi sollevati, citiamo la domanda se ci possa essere più di una ontologia fondazionale (nessuno dubita che un certo dominio possa essere concettualizzato in modo diverso, ma che dire quando si cerca di passare alle categorie più generali?) e il problema di come far comunicare agenti che utilizzano diverse concettualizzazioni del mondo.

Il lavoro di Enrico Franconi illustra come la logica (vista come strumento per la rappresentazione della conoscenza) giochi un ruolo fondamentale nel formalizzare e risolvere alcuni problemi fondamentali per il Semantic Web, come per esempio il problema di porre domande a un sistema in cui la conoscenza è distribuita su diverse “fonti” e quello di integrare i risultati che da una tale ricerca vengono restituiti.

Il lavoro di Serafini e Zanobini presenta una possibile soluzione alla domanda di come far comunicare agenti con diversi modelli del mondo (che utilizzano diverse ontologie). L'approccio presentato è fortemente motivato dal mondo del Semantic Web, dato che cerca di fare meno assunzioni possibili sull'esistenza di un accordo preliminare sul significato delle parole che utilizzano i diversi agenti coinvolti. Questo lavoro fa vedere come tecniche classiche di intelligenza artificiale, come per esempio il ragionamento automatico per la determinazione della soddisfacibilità di una formula logica, possano trovare una nuova e interessante applicazione nell'ambito del Semantic Web.

Roncaglia propone una riflessione sul problema di generare i metadati necessari per il funzionamento del Semantic Web. Il suo contributo presenta una prospettiva alquanto originale rispetto agli approcci comunemente adottati dai ricercatori impegnati nella costruzione del Semantic Web, in quanto individua nella blogosfera (vale a dire nello spazio condiviso degli utenti dei weblog) un ambiente ideale per la sperimentazione degli strumenti del Semantic Web, e questo perchè la blogosfera garantisce 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.

Infine, il contributo di Bonifacio solleva alcuni interessanti problemi che potremmo chiamare di *economia del significato*, ovvero problemi che concernono il legame – ancora poco esplorato, e non solo nella comunità del Semantic Web – tra conoscenza, organizzazione e valore. Bonifacio propone un embrione di modello economico per spiegare come e perché il significato delle parole evolve all'interno di comunità umane, e del valore economico che tale evoluzione può avere. Il suo contributo termina con una visione del Semantic Web inteso come “laboratorio della modernità”, una visione che ci pare una sfida molto interessante per i lettori della rivista.