

0 1 0 1 0 1 0 0 0 1
k n o w
w h a t
y o u
m e a n . . .

IL WEB SEMANTICO

Quando Internet diventa intelligente:

agenti software e rappresentazioni condivise

per un'automazione dei servizi in rete... e a casa vostra

Lo stereo suonava *We Can Work It Out* dei Beatles, quando squillò il telefono. Non appena Pete rispose, il telefono abbassò il volume della musica, inviando un messaggio al dispositivo di controllo. Era Lucy, la sorella, in linea dallo studio del medico: «La mamma ha bisogno di vedere uno specialista e di una serie di sedute di fisioterapia. Sto per ordinare al mio agente di fissare un appuntamento». In ufficio, Lucy istruì il suo agente Web semantico con il browser portatile. L'agente recuperò in tempo reale, dall'agente del medico, le informazioni sul trattamento, controllò l'elenco dei fisioterapisti e cercò quelli convenzionati con l'assicurazione della mamma, a meno di 30 chilometri di distanza da casa sua e con un giudizio di eccellente o molto buono nei servizi offerti. L'agente iniziò poi a cercare di far coincidere gli orari degli appuntamenti (forniti attraverso il sito Web di ciascuno studio) e le fitte agende di Pete e Lucy. In pochi minuti l'agente trovò una soluzione, ma a Pete non piacque: l'ospedale universitario era dall'altra parte della città rispetto alla casa della mamma, ed egli avrebbe dovuto tornare indietro proprio all'ora di punta. Istruì allora il suo agente in modo che ripettesse la ricerca, ma con indicazioni più precise in fatto di luoghi e di orari. L'agente di Lucy lo assistette nell'operazione, fornendogli le password di accesso ai siti visitati.

di
TIM BERNERS-LEE,
JAMES HENDLER
e ORA LASSILA

Illustrazioni
di MIGUEL SALMERON

L'agente presentò un nuovo piano, un fisioterapista molto più vicino e un orario più comodo: ma c'erano due note di avvertimento. Prima di tutto, Pete avrebbe dovuto spostare un paio dei suoi appuntamenti. Controllò di che cosa si trattava: nessun problema. L'altro punto era che l'agenzia assicurativa non comprendeva questo Centro nell'elenco dei fisioterapisti: «Il tipo di servizio e le convenzioni assicurative sono stati verificati in altro modo (dettagli?)». Lucy diede il suo assenso più o meno nello stesso momento in cui Pete borbottava «Risparmiami i dettagli», e l'appuntamento fu fissato.

Esprimere il significato

Pete e Lucy possono usare i loro agenti per svolgere tutti questi compiti non grazie al World Wide Web che conosciamo, ma a un Web semantico che ne sarà l'evoluzione futura. Il Web sta rivoluzionando le comunicazioni e il commercio, ma la maggior parte del suo contenuto è progettata per essere letta da esseri umani, e non per essere manipolato da programmi. I computer possono scorrere le pagine Web alla ricerca di schemi (qui c'è un titolo, qui un link, qui un'immagine) ma non hanno un sistema adeguato per trattare la semantica (questa è la homepage del Centro fisioterapico Hartmann e Strauss, questo link porta al *curriculum vitae* del dottor Hartmann, quel programma vi mostra l'elenco degli orari disponibili per gli appuntamenti).

Il Web semantico darà struttura al contenuto significativo della pagine Web, creando un ambiente dove gli agenti software possano svolgere velocemente compiti complessi per i loro utenti. Uno di questi agenti che arrivi alla pagina Web del Centro fisioterapico non saprà solo che quella pagina contiene parole chiave come «trattamento, medicina, fisico, terapia», ma anche che il dottor Hartmann effettua visite il lunedì, mercoledì e venerdì, e che il programma per gli appuntamenti accetta le date nel formato gg/mm/aaaa e presenta un elenco di orari disponibili. Per far questo non ci sarà bisogno di un'intelligenza artificiale. Il contenuto semantico sarà codificato nella pagina Web dal direttore amministrativo del Centro usando programmi *user-friendly* per scrivere pagine Web semantiche insieme con le risorse elencate nel sito dell'Associazione di fisioterapia.

Il Web semantico è un'estensione di quello attuale, in cui alle informazioni viene attribuito un significato definito, che permette a computer e utenti umani di lavorare meglio insieme. I primi passi per inserire questo Web semantico nella struttura di quello esistente sono già in corso. Presto questi sviluppi introdurranno nuove e significative funzionalità, via via che le macchine diventeranno sempre più abili a elaborare e «capire» i dati che al momento si limitano a mostrare.

La proprietà essenziale del World Wide Web è la sua universalità. La potenza dei link è data dal fatto che qualsiasi cosa può collegarsi a qualsiasi altra. La tecnologia del Web, quindi, non deve discriminare tra le brutte copie e i lavori finiti, fra informazioni commerciali e accademiche, o tra culture, lingue, media e così via. L'informazione varia lungo molteplici assi. Uno di questi è la differenza fra l'informazione prodotta per l'impiego da parte di utenti umani e quella prodotta per le macchine. A un'estremità della scala abbiamo di tutto, dai comunicati commerciali di cinque secondi fino a opere poetiche. Dall'altra parte abbiamo i database, i programmi e le misurazioni dei sensori. Finora, il Web è stato sviluppato soprattutto come veicolo per documenti destinati a utenti umani. Il Web semantico aspira a porre rimedio a questa situazione.

Come l'Internet attuale, anche il Web semantico sarà il più decentralizzato possibile. Simili sistemi suscitano enorme interesse, e i loro benefici sono difficili o impossibili da prevedere in anticipo. La decentralizzazione richiede però compromessi: il Web deve gettare via l'idea di una coerenza totale, sopportando il fami-



www.w3.org

Indirizzo Internet del World Wide Web Consortium (W3C).

www.w3.org/2001/sw/

Un sito che contiene informazioni approfondite sulle ricerche del W3C riguardo al Web semantico.

www.SemanticWeb.org/knowmarkup.html

Un'introduzione alle ontologie.

www.cs.umd.edu/projects/plus/SHOE/faq.html

Domande e risposte riguardo a semplici estensioni di HTML alle ontologie.

www.daml.org/

Home page del DARPA Agent Markup Language (DAML).

IN PILLOLE

- Fino a oggi, il World Wide Web si è sviluppato più rapidamente come strumento di contenuti per uso «umano» piuttosto che come mezzo di informazione per l'elaborazione automatizzata. Migliorando il contenuto delle pagine Web con dati specificamente progettati per i computer e aggiungendo documenti comprensibili alle macchine, trasformeremo il Web di oggi nel Web semantico.
- I computer troveranno il significato dei dati semantici seguendo link ipertestuali verso definizioni di parole chiave e regole logiche per combinarle. L'infrastruttura che ne risulterà permetterà lo sviluppo di servizi Web automatizzati grazie all'uso di agenti software di grande efficacia.
- Gli utenti comporranno pagine Web semantiche e aggiungeranno definizioni e regole utilizzando software in grado di attribuire i significati e di usare regole logiche definite.

gerato messaggio «Error 404: Not Found», ma permettendo una crescita esponenziale senza limitazioni. Per essere universale, cioè per consentire un facile accesso a qualsiasi informazione, il Web semantico deve essere il meno costrittivo possibile.

La rappresentazione della conoscenza

Affinché il Web semantico funzioni, i computer devono avere accesso a serie strutturate di informazioni e a regole di deduzione che permettano un ragionamento automatizzato. Chi lavora sull'intelligenza artificiale studia simili sistemi da molto tempo prima che venisse sviluppato il Web. Questo settore, spesso chiamato rappresentazione della conoscenza, è oggi in uno stato di sviluppo comparabile con quello della tecnologia degli ipertesti prima dell'avvento del Web: è chiaramente una buona idea ed esistono alcune dimostrazioni interessanti, ma non ha ancora cambiato il mondo. L'idea contiene il seme di importanti applicazioni, ma deve essere collegata in un unico sistema globale per poter liberare la sua piena potenza.

I sistemi tradizionali di rappresentazione della conoscenza sono centralizzati e richiedono di condividere la stessa definizione di concetti comuni. Ma un controllo centrale è opprimente, e con la crescita delle dimensioni e degli obiettivi, un simile sistema diventa rapidamente ingestibile. Questi sistemi di solito

GLOSSARIO

► **HTML: Hypertext Markup Language.** È il linguaggio usato per codificare tutte le caratteristiche delle pagine Web. Fa uso di «comandi» standard come <H1> e <BODY>, il cui significato e interpretazione sono univocamente definiti dal World Wide Web Consortium.

► **XML: eXtensible Markup Language.** È un linguaggio come HTML che permette di definire comandi personalizzati, ma non ha un meccanismo interno che permetta di comunicare ad altri utenti il significato di tali comandi.

► **Resource:** In gergo, una qualsiasi entità: pagine Web, parti di esse, dispositivi, persone e così via.

► **URL: Uniform Resource Locator.** I codici (come <http://www.lescienze.it>) usati nei link ipertestuali.

► **URI: Universal Resource Identifier.** Gli URL sono il tipo più familiare di URI. Un URI definisce un'entità, senza necessariamente richiamare la sua locazione sul Web.

► **RDF: Resource Description Framework.** Uno schema per definire l'informazione. RDF fornisce la tecnologia per esprimere il significato di termini e concetti in una forma che i computer possano elaborare. RDF può utilizzare il linguaggio XML per la sua sintassi e gli URI per specificare entità, concetti, proprietà e relazioni.

► **Ontologie:** Collezioni di frasi, scritte in un linguaggio come RDF, che definiscono le relazioni tra concetti e specificano le regole logiche. I computer «capiranno» il senso dei dati su una pagina Web seguendo i link a determinate ontologie.

► **Agente:** Software che esegue compiti definiti da un utente senza il diretto controllo da parte dell'utente stesso. Gli agenti raccolgono, filtrano ed elaborano l'informazione che si trova sul Web.



LE RICERCHE SUL WEB finiscono spesso in un'interminabile serie di risultati del tutto irrilevanti che devono essere filtrati dall'utente. Se cercate il termine inglese *cook* (cuoco, cucinare), il computer non può sapere se cercate uno chef, una ricetta, le Isole Cook o qualcuno che, sfortunatamente, si chiama proprio così... Il problema è che, cercando *cook*, non si fornisce alcun significato, o contenuto semantico, al computer, che non può discernere tra le diverse opzioni.

sono anche limitati dal punto di vista delle domande a cui un computer può rispondere. Il problema ricorda il teorema di Gödel: ogni sistema abbastanza complesso da essere utile contiene qualche domanda a cui non c'è risposta, proprio come una sofisticata versione del semplice paradosso: «Questa affermazione è falsa». Per evitare simili problemi, ogni sistema di rappresentazione della conoscenza ha una propria serie di regole per fare deduzioni sui dati. Per esempio, un sistema che lavora su un database di alberi genealogici potrebbe avere la regola: «la sorella di una madre è una zia». Anche se i dati potrebbero essere trasferiti da un sistema a un altro, non si potrebbe fare altrettanto con la regola.

I ricercatori che studiano il Web semantico, al contrario, accettano che paradossi e domande senza risposta siano il prezzo che si deve pagare per raggiungere la versatilità. Noi rendiamo il linguaggio delle regole espressivo quanto è necessario per consentire al Web di ragionare nel modo più ampio possibile. Questa filosofia è simile a quella del Web convenzionale. All'inizio del World Wide Web, i suoi detrattori affermarono che non sarebbe mai diventato una biblioteca ben organizzata: senza un database centrale e una struttura ad albero non si sarebbe mai stati sicuri di trovare tutto. E avevano ragione. Ma la potenza del sistema rese disponibili enormi quantità di informazioni, e oggi i motori di ricerca producono indici soddisfacenti del materiale in rete.

La sfida del Web semantico, perciò, è quella di fornire un linguaggio che esprima sia i dati sia le regole per ragionare su di essi, e che permetta di esportare in rete le regole di qualsiasi sistema di rappresentazione della conoscenza. Le regole di un sistema probabilmente non potranno essere reimportate dal Web in un altro sistema, ma questo non è un grave difetto. Il punto è quello di rendere disponibile l'informazione.

Aggiungere logica al Web è il problema che si pone in questo momento alla comunità del Web semantico, compito reso complicato da una serie di dilemmi matematici e tecnologici. La logica deve essere abbastanza potente da descrivere, per esempio, le proprietà di proprietà, ma non così potente da generare paradossi. Fortunatamente, la maggioranza delle informazioni che vogliamo esprimere è nella forma «un dado a testa esagonale è un tipo di bullone per macchina», che può essere facilmente scritta nei linguaggi esistenti, con l'aiuto di un vocabolario.

Due importanti tecnologie per lo sviluppo del Web semantico sono già disponibili: l'*eXtensible Markup Language* (XML) e il *Resource Description Framework* (RDF). Il linguaggio XML permette di creare *tag* personalizzati (etichette che commentano pagine Web o porzioni di testo su una pagina). I programmi possono fare uso di queste etichette in modi sofisticati, ma il programmatore che li scrive deve sapere a che scopo chi ha scritto la pagina Web utilizza un *tag*. In breve, XML permette agli utenti di aggiungere una struttura ai loro documenti, ma non dice nulla su che cosa significhi.

Il significato è espresso da RDF che lo codifica in serie di triple, in cui ogni tripla è analoga a soggetto, verbo e complemento oggetto di una frase elementare. In RDF, un documento fa affermazioni sul fatto che cose particolari (persone, pagine Web o altro) hanno proprietà (come «è la sorella di», «è l'autore di») con certi valori (un'altra persona, un'altra pagina Web). Le triple sono un modo naturale per descrivere gran parte dei dati elaborati dalle macchine. Soggetto e complemento oggetto sono identificati da un *Universal Resource Identifier* (URI, o identificatore universale di risorse), analogo a quello utilizzato da un link a una pagina Web (gli URL, *Uniform Resource Locator*, sono i più comuni tipi di URI). Anche i verbi sono identificati da URI, che permettono a chiunque di definire un nuovo concetto, un nuovo verbo, semplicemente definendo un loro URI da qualche parte sul Web.

Spesso il linguaggio umano usa lo stesso termine per esprimere cose un po' diverse, ma l'automazione non può farlo. Immaginate di assumere un servizio di clown messenger per con-

segnare palloncini ai clienti in occasione del loro compleanno. Sfortunatamente il servizio preleva gli indirizzi da un database, non sapendo che gli «indirizzi» indicati sono quelli dove vengono mandate le fatture, e molti di essi sono solo caselle postali. I clown finirebbero con l'intrattenere un certo numero di impiegati delle poste, il che non è necessariamente una brutta cosa, ma non è ciò che si desiderava. Il problema si risolve usando un diverso URI per ciascun concetto. Un indirizzo che è un luogo può essere così differenziato da una semplice casella postale.

Le triple di RDF formano ragnatele di informazioni riguardo a cose correlate (o «risorse», nel gergo del Web). Poiché RDF usa XML per codificare l'informazione in un documento, gli URI garantiscono che i concetti non siano solo parole in un documento, ma che siano legati a una definizione univoca che chiunque può trovare in rete. Immaginiamo di avere accesso a una varietà di database con informazioni su persone, compreso il loro indirizzo. Se vogliamo trovare persone che vivono in una zona con uno specifico codice di avviamento postale, dobbiamo sapere quali campi di ogni database rappresentano i nomi e quali rappresentano i codici postali. RDF può specificare che «(il campo 5 nel database A) (è un campo di tipo) (codice di avviamento postale)», e così via, usando per ciascun termine URI anziché frasi.

Ontologie in rete

I guai non finiscono qui: due database possono usare identificatori diversi per lo stesso concetto. Un programma che vuole confrontare o combinare informazioni da due database deve sapere quali termini vengono usati per indicare la stessa cosa. Idealmente, il programma deve avere un mezzo per scoprire tali significati comuni per qualsiasi database.

Una soluzione è data dal terzo componente del Web semantico: raccolte di informazioni che chiamiamo ontologie. In filosofia, un'ontologia è una teoria sulla natura dell'esistenza, su quali tipi di cose esistono. I ricercatori del Web e dell'intelligenza artificiale hanno adottato questo termine per il loro gergo, e per loro l'ontologia è un documento o un file che definisce formalmente le relazioni fra termini. Il più caratteristico tipo di ontologia per il Web ha una tassonomia e una serie di regole di deduzione.

La tassonomia definisce classi di oggetti e le loro relazioni. Per esempio, un indirizzo può essere definito come un tipo di ubicazione, e il codice cittadino può venire definito in modo da essere applicato solo a ubicazioni e così via. Le classi, le sotto-

E qual è la killer app?

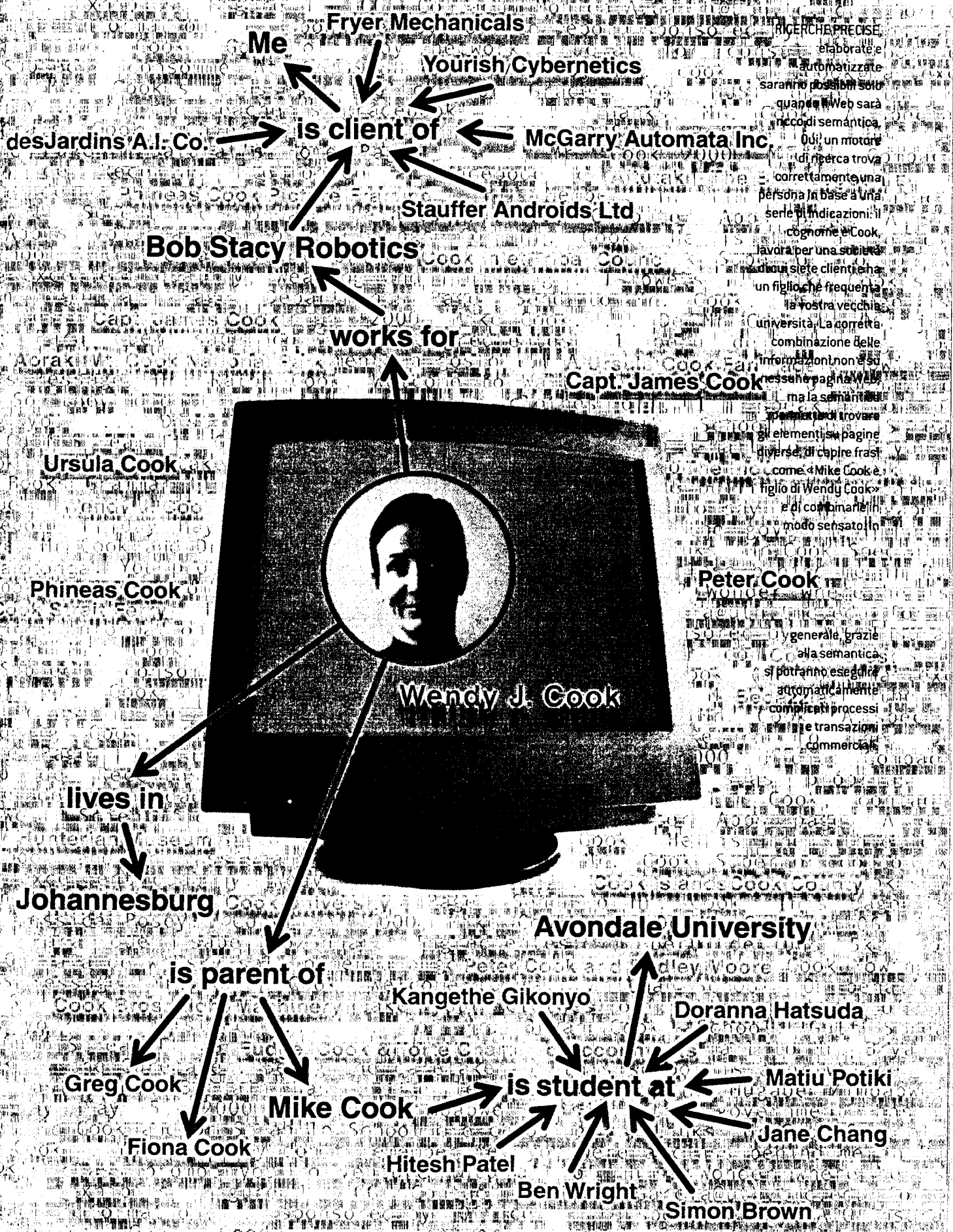
Dopo una presentazione del Web semantico, spesso ci sentiamo chiedere: «Bene, e quale sarebbe la killer application del Web semantico?». La killer app di una tecnologia è l'applicazione che porta un utente a provarla e poi a usarla con assiduità. La radio a transistor fu la killer app dei transistor, e il telefono cellulare è quella della tecnologia senza fili.

«Il Web semantico è già la killer app?» rispondiamo noi. Di solito ci scambiano per matti, così tocca a noi fare una domanda: «Bene, il World Wide Web è la killer app di che cosa?». Di solito ci troviamo di fronte gente dagli occhi sgranati, e allora rispondiamo noi stessi: «Se il Web è la killer app di Internet, il Web semantico è una killer app della stessa grandezza».

Il punto è che le capacità del Web semantico sono troppo generali per essere pensate solo in funzione della soluzione di un particolare problema o per creare qualche nuovo aggeggiato.

Avrà applicazioni che non abbiamo mai nemmeno sognato. Nondimeno, possiamo intravedere qualcuna delle disarmanti applicazioni iniziali. I cataloghi on line avranno contrassegni semantici utili sia agli acquirenti sia ai venditori.

Sarà più facile eseguire piccole transazioni commerciali con meccanismi di sicurezza e in autonomia. E infine, si potranno fare tranquillamente prenotazioni per un lungo viaggio all'estero: voli, alberghi e perfino un concerto o un biglietto di teatro. Tutte le prenotazioni si caricheranno automaticamente nella vostra agenda e le spese saranno addebitate direttamente sul vostro conto corrente, qualunque sia il software semantico che utilizzate. Niente più copia e incolla dall'e-mail. E non ci sarà più bisogno, per l'e-commerce, di inserire dati in una mezza dozzina di formati diversi o di creare un formato a parte. Beh, vi pare poco?



RICERCHES PRECISE,
elaborate e
robotizzate
saranno possibili
quando Web sarà
ricco di semantica.
Où: un motore
(di ricerca trova
correttamente una
persona in base a una
serie di indicazioni:
il cognome e Cook,
lavora per una società
e un siete clienti e ha
un figlio che frequenta
la vostra vecchia
università. La corretta
combinazione delle
informazioni non è
nessa pagina Web
ma la semantica
permette di trovare
gli elementi su pagine
diverse. Il capire frasi
come «Mike Cook è
figlio di Wendy Cook»
e di combinarli in
modo sensato in
un unico documento.
Peter Cook è
il generale, grazie
alla semantica
si potranno eseguire
automaticamente
complicati processi
e transazioni
commerciali.

classi e le relazioni fra enti sono uno strumento potente per l'uso del Web. Possiamo esprimere un gran numero di relazioni fra enti assegnando proprietà alle classi e permettendo alle sottoclassi di «ereditare» tali proprietà. Se il codice cittadino deve essere del tipo «città», possiamo trattare il sito Web associato a un codice cittadino anche se nessun database connette direttamente questo codice con un sito Web.

Le regole di deduzione danno potenza aggiuntiva. Un'ontologia può esprimere la regola: «se un codice cittadino è associato a un codice statale e un codice indirizzo utilizza il codice cittadino, allora quell'indirizzo ha il relativo codice statale». Un programma può facilmente dedurre, per esempio, che un indirizzo della Cornell University di Ithaca deve essere nello Stato di New York, che si trova negli Stati Uniti, e quindi deve essere scritto secondo gli standard americani. Il computer non «capisce» nessuna di queste informazioni, ma può manipolare i termini molto efficacemente e in modi significativi per l'utente.

Quando in rete si trovano pagine ontologiche, iniziano a delinearsi le soluzioni ai problemi di terminologia. Il significato di termini o di codici XML utilizzati su una pagina Web può essere definito da puntatori che conducono all'ontologia. Gli stessi problemi di prima sorgono se io definisco un puntatore a un'ontologia che definisce gli indirizzi che contengono un codice di avviamento e voi puntate a un'altra che usa un codice postale. Questo tipo di confusione potrebbe essere risolto se le ontologie fornissero le relazioni di equivalenza: una o entrambe le nostre ontologie potrebbero contenere informazioni sul fatto che i due codici sono la stessa cosa e il problema sarebbe risolto.

Il problema dell'invio dei clown si risolve quando i due database puntano a due differenti definizioni di indirizzo. Il programma viene avvertito che devono prima essere risolte le differenze fra i termini. La soluzione potrebbe includere l'uso di un servizio che prende un elenco di indirizzi postali e lo converte in uno di indirizzi fisici, riconoscendo e rimuovendo le caselle postali e altri indirizzi inadatti. La struttura e la semantica fornite dalle ontologie rendono più semplice gestire un simile servizio e fanno sì che il suo impiego sia completamente trasparente.

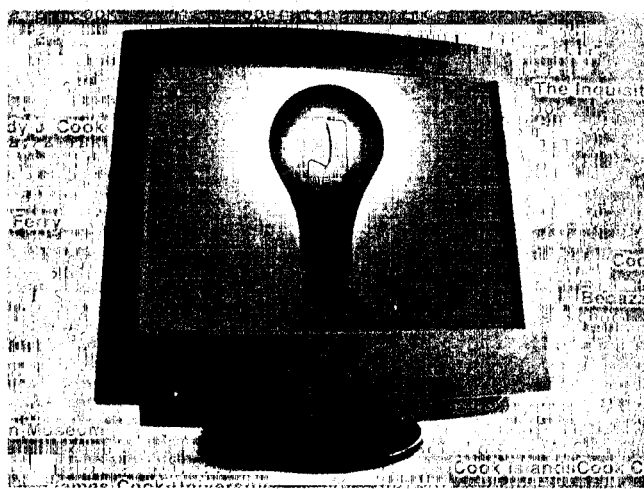
Le ontologie possono migliorare il funzionamento del Web in molti modi. Per esempio, esse possono essere usate in modo semplice per migliorare l'accuratezza delle ricerche in rete. Supponete di voler acquistare un nuovo cancello per la vostra casa e di fare una ricerca con la parola chiave *gates*. Troverete molte più pagine su Bill Gates o sull'attrice Gates McFadden o altre persone e luoghi chiamati Gates, piuttosto che su accessori per la casa. Se le pagine Web contengono puntatori alle ontologie, il problema può essere alleviato: potreste cercare informazioni usando un programma per selezionare una particolare definizione di *gate*.

La vera potenza delle ontologie si scatena quando i computer ne fanno un uso più sofisticato per correlare le informazioni su una pagina con le relative strutture di conoscenza e regole di deduzione. Per esempio, se puntate il vostro *browser* Web alla pagina <http://www.cs.umd.edu/~hendler>, vedrete la normale pagina Web del «Dr. James Hendler». Non vi è difficile trovare il collegamento a una breve nota biografica e leggere che Hendler ha conseguito il suo Ph.D. alla Brown University. Un programma per computer che tentasse di trovare la stessa informazione, tuttavia, dovrebbe essere molto complesso per capire che questa informazione potrebbe trovarsi in una biografia e per capire i termini inglesi in essa usati.

Per il computer, la pagina Web è collegata a una pagina ontologica che definisce le informazioni a proposito dei dipartimenti di informatica. Per esempio, i professori lavorano nelle Università, generalmente hanno il dottorato e così via. Un'ulteriore etichettatura della pagina (non mostrata da un browser

normale) usa i concetti propri dell'ontologia per specificare che Hendler ricevette il suo Ph.D. dall'ente descritto dall'URI <http://www.brown.edu/>: la pagina Web della Brown University. Il computer può anche trovare che Hendler è membro di un particolare progetto di ricerca, ha un certo indirizzo e-mail e così via. Tutte queste informazioni sono facilmente elaborate da un computer e possono essere usate per rispondere a domande (dove ha conseguito il dottorato il dottor Hendler?) che al momento richiederebbero un essere umano per analizzare il contenuto delle pagine trovate dal motore di ricerca.

Inoltre, questa etichettatura rende molto più semplice mettere a punto programmi che possano affrontare domande complicate, la cui risposta non risiede in una singola pagina Web. Supponete di voler trovare la signora Cook che avete incontrato l'anno scorso a un congresso. Non rammentate il suo nome di battesimo, ma ricordate che lavorava per uno dei vostri clienti e che suo figlio studiava nella vostra stessa Università. Un programma di ricerca intelligente può scorrere tutte le pagine delle persone il cui nome è Cook (saltando le pagine che parlano di cuochi - *cook* in inglese - dell'Isola di Cook e così



via); può trovare quelle che accennano al fatto di lavorare per una società che è nel vostro elenco di clienti e seguire i collegamenti alle pagine Web dei loro figli per trovare quelli che stanno studiando nel posto giusto.

Agenti software per il Web semantico

La vera potenza del Web semantico si realizzerà quando verranno creati molti programmi che raccolgano il contenuto Web da diverse fonti, elaborino le informazioni e scambino i risultati con altri programmi. La potenza di questi «agenti» software crescerà esponenzialmente via via che saranno disponibili sempre più pagine Web leggibili da macchine e servizi automatici. Il Web semantico promuove questa cooperazione: anche agenti che non erano progettati per lavorare insieme possono scambiarsi dati, se questi sono dotati di una semantica.

Una caratteristica importante del funzionamento degli agenti sarà lo scambio di «prove» scritte nel linguaggio unificante del Web semantico (il linguaggio che esprime le deduzioni logiche fatte usando regole e informazioni come quelle specificate nelle ontologie). Supponiamo che un servizio on line abbia localizzato l'informazione per mettervi in contatto con la signora Cook: con vostra sorpresa, la colloca a Johannesburg. Naturalmente volete verificare l'informazione, e quindi il vostro computer chiede al servizio una prova di questa risposta, che viene fornita traducendo il suo ragionamento interno nel linguaggio unificante del Web semantico. Un motore di deduzione nel vostro

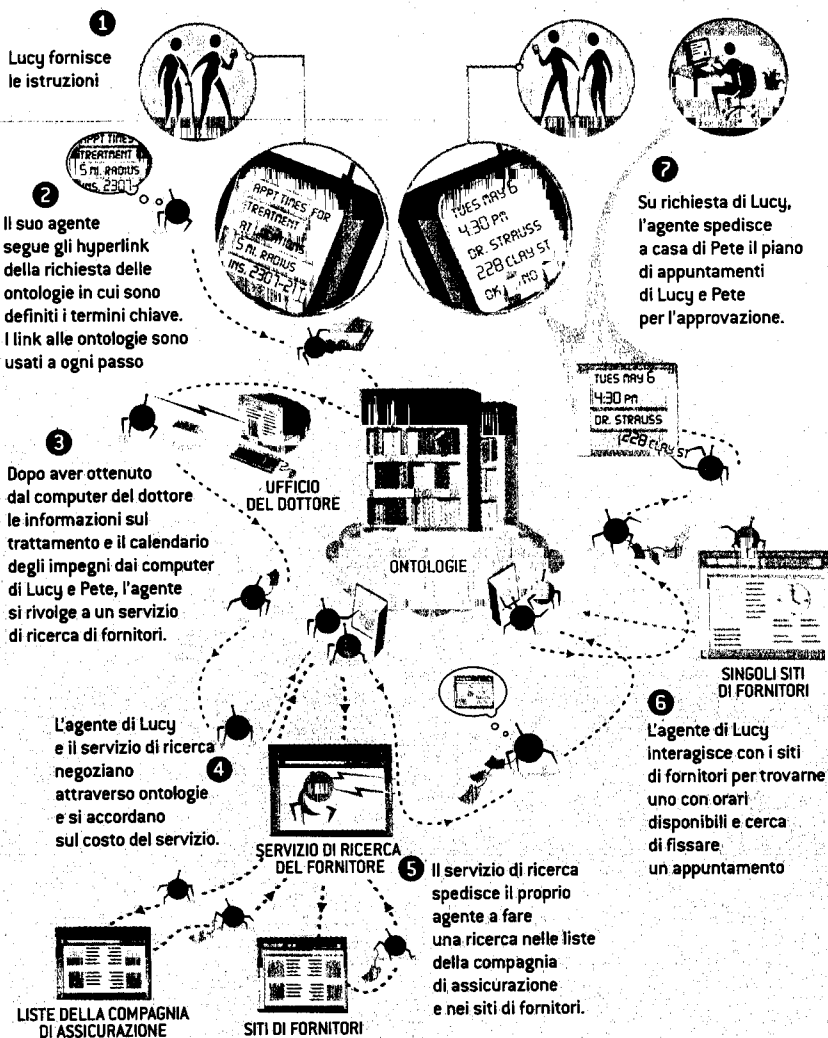
GLI AGENTI SOFTWARE avranno vita facile, con più contenuto semantico sul Web. Nell'esempio, l'agente di Lucy rintraccia una clinica fisioterapica per la madre combinando una serie di criteri di scelta e fissa appuntamenti in accordo con gli impegni suoi e del fratello, Pete. Le ontologie che definiscono il significato semantico dei dati permettono all'agente di capire i contenuti, di interagire con i siti e di usare servizi automatizzati.

GLI AUTORI

BERNERS-LEE dirige il World Wide Web Consortium (W3C) ed è ricercatore presso il Laboratory for Computer Science del MIT. Quando inventò il Web, nel 1989, intendeva introdurre più semantica di quanta ve ne sia. Nel suo libro *Weaving the Web* definisce il Web semantico e ciò che occorrerà per realizzarlo.

HENDLER insegna informatica all'Università del Maryland, dove studia la rappresentazione della conoscenza in rete, e ha sviluppato SHOE, il primo linguaggio di rappresentazione basato sul Web, per dimostrare le capacità degli agenti software. È responsabile di un progetto di ricerca della Defense Advanced Research Projects Agency di Arlington.

LASSILA è ricercatore al Centro ricerche Nokia di Boston. Frustrato dalla difficoltà di costruire agenti e di automatizzare i compiti sul Web, è stato coautore delle specifiche di RDF del W3C, che fungono da fondamenta per molti lavori tesi alla creazione del Web semantico.



computer può facilmente verificare che la signora Cook è effettivamente quella che stavate cercando, e vi può mostrare le pagine Web più rilevanti, se avete ancora dubbi. Sebbene siano ancora lontani dal sondare le profondità del potenziale del Web semantico, alcuni programmi possono già scambiare prove, usando la versione preliminare del linguaggio unificante.

Un'altra caratteristica vitale saranno le firme elettroniche: blocchi di dati crittografati che computer e agenti possono usare per verificare che le informazioni allegate siano state fornite da una fonte di fiducia. Ovviamente volete essere sicuri che un messaggio inviato al vostro programma di contabilità sul fatto che avete un debito verso un venditore on line non è una falsa informazione generata dall'adolescente maniaco di computer della porta accanto. Gli agenti dovrebbero essere scettici a proposito delle affermazioni che leggono sul Web semantico fino a quando non hanno verificato la loro fonte.

Molti servizi automatizzati basati sul Web esistono già senza semantica ma altri programmi, come gli agenti, non hanno modo di localizzare un servizio che esegua una funzione specifica. Questo processo - il ritrovamento di un servizio - può avvenire solo quando c'è un linguaggio comune per descrivere il servizio in un modo che permetta agli agenti di capire la funzionalità offerta e soprattutto come servirsene. Servizi e agenti possono pubblicizzare la loro funzione, per esempio, depositando descrizioni in elenchi simili alle Pagine gialle: una forma di pubblicità molto migliore dei banner che infestano numerosi siti Web odierni.

Alcuni metodi di riconoscimento dei servizi esistono già, co-

me Universal Plug and Play, della Microsoft, concentrato sulla connessione di diversi tipi di dispositivi hardware, e Jini, della Sun Microsystems, che ha invece lo scopo di connettere servizi. Queste iniziative, tuttavia, affrontano il problema a livello strutturale o sintattico e si basano pesantemente sulla standardizzazione di una serie predeterminata di descrizioni di funzionalità. La standardizzazione può spingersi solo fino a un certo limite, poiché non possiamo anticipare tutti i possibili bisogni futuri.

Il Web semantico è più flessibile. Gli agenti dei produttori e dei consumatori possono intendersi a vicenda scambiandosi ontologie, che forniscono il vocabolario necessario alla discussione. Gli agenti possono anche sviluppare nuove capacità di ragionamento quando scoprono nuove ontologie. Le semantiche rendono anche più semplice sfruttare un servizio che risponde solo parzialmente alla richiesta.

Un processo tipico implicherà la creazione di una «catena di valori» in cui sottoinsiemi di informazione vengono passati da un agente all'altro, e ognuno di essi «aggiunge un valore» per costruire il prodotto finale richiesto dall'utente. Beninteso: per creare automaticamente sofisticate catene di valori su richiesta, alcuni agenti dovranno usare tecniche di intelligenza artificiale in aggiunta a quelle offerte dal Web semantico. Ma quest'ultimo fornirà l'architettura per realizzare queste tecnologie.

L'insieme di tutte queste caratteristiche produce le capacità messe in mostra dagli agenti di Pete e Lucy nell'episodio che apre l'articolo. I loro agenti potrebbero delegare parte del compito ad altri servizi o ad agenti scoperti attraverso la pubblicità.

Per esempio, gli agenti potrebbero avere un servizio sicuro per procurarsi un elenco di fornitori e decidere quali siano conformi a specifiche convenzioni assicurative e modalità di trattamento. L'elenco dei fornitori sarebbe a sua volta fornito da un altro servizio di ricerca, e così via. Queste attività producono catene in cui grandi quantità di dati distribuite in tutto il Web (e quasi inutilizzabili in quella forma) sono via via ridotte ai pochi dati importanti per Pete e Lucy: una serie di appuntamenti che si adattano ai loro impegni e alle loro richieste.

Il passo successivo per il Web semantico sarà quello di uscire dal regno virtuale ed estendersi nel mondo fisico. Gli URI possono puntare a qualsiasi cosa, compresi enti fisici, il che significa che si può usare il linguaggio RDF per descrivere dispositivi come telefoni cellulari e televisori. Questi dispositivi possono allora dichiarare le loro funzionalità - ciò che sono in grado di fare e come sono controllati - in modo simile agli agenti software. E dato che essi sono molto più flessibili di sistemi come Plug and Play, un simile approccio semantico apre un mondo di enormi possibilità.

Per esempio, ciò che oggi è chiamato «automazione domestica» richiede un'accurata configurazione che permetta a tutti gli elettrodomestici di funzionare insieme. Le descrizioni semanti-

che delle capacità e delle funzionalità dei diversi apparecchi ci permetteranno di conseguire l'automazione con un intervento umano minimo. Un esempio banale è il momento in cui Pete risponde al telefono e il volume dell'impianto stereo si abbassa automaticamente. Invece di programmare ogni specifico elettrodomestico, Pete definisce una volta per tutte questa funzione per comandare ogni dispositivo locale dotato di un controllo del volume, come il televisore, il DVD e anche i programmi multimediali sul computer portatile.

Il primo passo concreto in questo settore è già stato intrapreso, con lavori per sviluppare uno standard che descriva le capacità funzionali dei dispositivi (come le dimensioni degli schermi) e le preferenze dell'utente. Costruito in RDF, questo standard si chiama *Composite Capability/Preference Profile* (CC/PP). All'inizio CC/PP permetterà ai telefoni cellulari e ad altri clienti Web non standard di descrivere le proprie caratteristiche in modo che il contenuto del Web possa essere adattato a loro in tempo reale. Più tardi, quando si aggiungerà la potenza dei linguaggi che gestiscono le ontologie e la logica, i dispositivi potrebbero automaticamente cercare e utilizzare servizi e altri dispositivi per ottenere informazioni e sviluppare funzionalità aggiuntive. Non è difficile immaginare un forno a microonde collegato al Web che consulta i produttori di cibo congelato per conoscere i migliori parametri di cottura.

Verso un'evoluzione della conoscenza

Il Web semantico non è solamente uno strumento per svolgere i singoli compiti che abbiamo discusso finora: se sarà progettato appropriatamente, potrà facilitare l'evoluzione dell'intera conoscenza umana.

Le attività dell'uomo sono caratterizzate da un'eterna tensione tra gruppi che agiscono indipendentemente e la necessità di inserirsi in una comunità più vasta. Un piccolo gruppo può sviluppare innovazioni velocemente e con efficienza, ma così facendo genera una sottocultura a cui concetti non sono capiti dagli altri. Coordinare le azioni di un vasto gruppo, tuttavia, è un processo lento, e richiede un'enorme mole di comunicazione. Il mondo funziona a tutti i livelli tra questi estremi, con una tendenza a iniziare dal piccolo - dalle idee personali - per muoversi col tempo verso una comprensione più generale.

Un processo essenziale è l'unione delle sottoculture quando diventa necessario un linguaggio comune. Spesso due gruppi sviluppano indipendentemente concetti molto simili e descrivere le relazioni tra essi è molto utile. Come un dizionario italiano/inglese, o una tabella di conversione di pesi e misure, le relazioni permettono di comunicare e di collaborare anche quando la condivisione dei concetti non ha ancora portato a una condivisione di termini.

Il Web semantico, definendo ogni concetto semplicemente con un URI, permette a chiunque di inventarne di nuovi, a suo piacimento. Il suo linguaggio logico di unificazione permetterà a questi concetti di essere progressivamente collegati in un Web universale. Ciò renderà disponibili conoscenza e lavoro umano all'analisi da parte degli agenti software, offrendoci una nuova classe di strumenti con i quali potremo vivere, lavorare e imparare insieme.

BIBLIOGRAFIA

BERNERS-LEE TIM e FISCHETTI MARK, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*, Harper San Francisco, 1999.

Una versione più estesa di questo articolo si trova sul sito di Scientific American (www.sciam.com) con materiale di approfondimento e link.

**Avete mai
navigato in internet,
aperto un file,
ascoltato un CD?**

Allora avete usato, magari
senza saperlo, semiconduttori,
fibre ottiche, polimeri avanzati
e gli altri materiali di nuova
generazione che stanno
cambiando il mondo
in cui viviamo.

Per essere protagonisti
nello studio,
nella progettazione
e nella produzione dei
materiali del futuro

**Laurea in
Scienza dei Materiali
Università di Milano Bicocca**

per informazioni:
tel. 02 6448 - 5137/5102
<http://www.mater.unimib.it/CdL>