

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

THE SEMANTIC WEB

A new form of Web content
that is meaningful to computers
will unleash a revolution of new possibilities

by
TIM BERNERS-LEE,
JAMES HENDLER and
ORA LASSILA

PHOTOILLUSTRATIONS BY MIGUEL SALMERON

*The entertainment system was belting out the Beatles' "We Can Work It Out" when the phone rang. When Pete answered, his phone turned the sound down by sending a message to all the other **local** devices that had a **volume control**. His sister, Lucy, was on the line from the doctor's office: "Mom needs to see a specialist and then has to have*

a series of physical therapy sessions. Bi-weekly or something. I'm going to have my agent set up the appointments." Pete immediately agreed to share the chauffeuring.

At the doctor's office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved information about Mom's **prescribed treatment** from the doctor's agent, looked up several lists of **providers**, and checked for the ones **in-plan** for Mom's insurance within a **20-mile radius** of her **home** and with a **rating** of **excellent** or **very good** on trusted rating services. It then began trying to find a match between available **appointment times** (supplied by the agents of individual providers through their Web sites) and Pete's and Lucy's busy schedules. (The emphasized keywords indicate terms whose semantics, or meaning, were defined for the agent through the Semantic Web.)

In a few minutes the agent presented them with a plan. Pete didn't like it—University Hospital was all the way across town from Mom's place, and he'd be driving back in the middle of rush hour. He set his own agent to redo the search with stricter preferences about **location** and **time**. Lucy's agent, having **complete trust** in Pete's agent in the context of the present task, automatically assisted by supplying access certificates and shortcuts to the data it had already sorted through.

Almost instantly the new plan was presented: a much closer clinic and earlier times—but there were two warning notes. First, Pete would have to reschedule a couple of his **less important** appointments. He checked what they were—not a problem. The other was something about the insurance company's list failing to include this provider under **physical therapists**: "Service type and insurance plan

status securely verified by other means," the agent reassured him. "(Details?)"

Lucy registered her assent at about the same moment Pete was muttering, "Spare me the details," and it was all set. (Of course, Pete couldn't resist the details and later that night had his agent explain how it had found that provider even though it wasn't on the proper list.)

Expressing Meaning

PETE AND LUCY could use their agents to carry out all these tasks thanks not to the World Wide Web of today but rather the Semantic Web that it will evolve into tomorrow. Most of the Web's content today is designed for humans to read, not for computer programs to manipulate meaningfully. Computers can adeptly parse Web pages for layout and routine processing—here a header, there a link to another page—but in general, computers have no reliable way to process the semantics: this is the home page of the Hartman and Strauss Physio Clinic, this link goes to Dr. Hartman's curriculum vitae.

The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users. Such an agent coming to the clinic's Web page will know not just that the page has keywords such as "treatment, medicine, physical, therapy"

Overview / *Semantic Web*

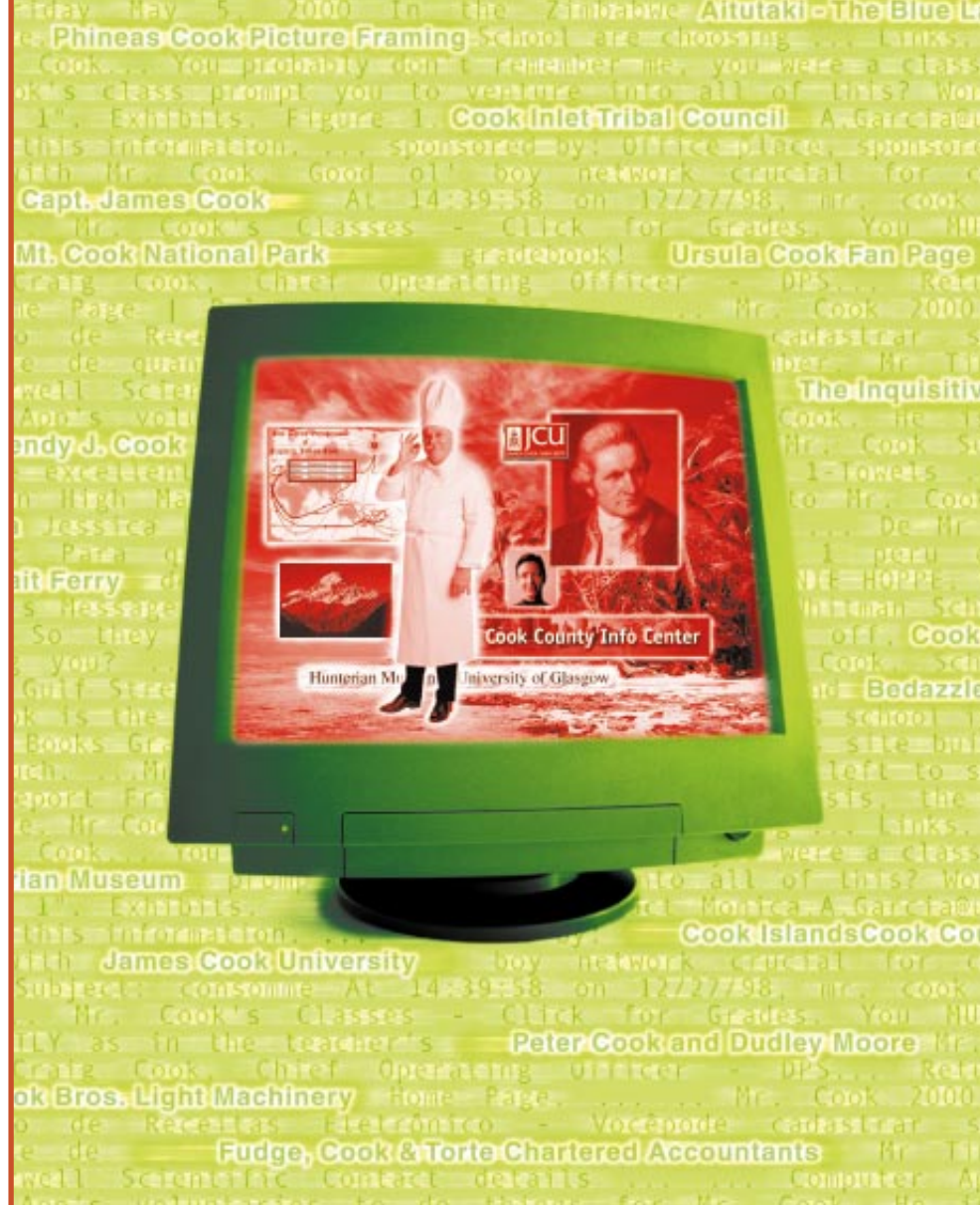
- To date, the World Wide Web has developed most rapidly as a medium of documents for people rather than of information that can be manipulated automatically. By augmenting Web pages with data targeted at computers and by adding documents solely for computers, we will transform the Web into the Semantic Web.
- Computers will find the meaning of semantic data by following hyperlinks to definitions of key terms and rules for reasoning about them logically. The resulting infrastructure will spur the development of automated Web services such as highly functional agents.
- Ordinary users will compose Semantic Web pages and add new definitions and rules using off-the-shelf software that will assist with semantic markup.

(as might be encoded today) but also that Dr. Hartman **works** at this **clinic** on **Mondays, Wednesdays** and **Fridays** and that the script takes a **date range** in **yyyy-mm-dd format** and returns **appointment times**. And it will “know” all this without needing artificial intelligence on the scale of 2001’s Hal or *Star Wars*’s C-3PO. Instead these semantics were encoded into the Web page when the clinic’s office manager (who never took Comp Sci 101) massaged it into shape using off-the-shelf software for writing Semantic Web pages along with resources listed on the Physical Therapy Association’s site.

The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The first steps in weaving the Semantic Web into the structure of the existing Web are already under way. In the near future, these developments will usher in significant new functionality as machines become much better able to process and “understand” the data that they merely display at present.

The essential property of the World Wide Web is its universality. The power of a hypertext link is that “anything can link to anything.” Web technology, therefore, must not discriminate between the scribbled draft and the polished performance, between commercial and academic information, or among cultures, languages, media and so on. Information varies along many axes. One of these is the difference between information produced primarily for human consumption and that produced mainly for machines. At one end of the scale we have everything from the five-second TV commercial to poetry. At the other end we have databases, programs and sensor output. To date, the Web has developed most rapidly as a medium of documents for people rather than for data and information that can be processed automatically. The Semantic Web aims to make up for this.

Like the Internet, the Semantic Web will be as decentralized as possible. Such Web-like systems generate a lot of excitement at every level, from major corporation to individual user, and provide bene-



WEB SEARCHES TODAY typically turn up innumerable completely irrelevant “hits,” requiring much manual filtering by the user. If you search using the keyword “cook,” for example, the computer has no way of knowing whether you are looking for a chef, information about how to cook something, or simply a place, person, business or some other entity with “cook” in its name. The problem is that the word “cook” has no meaning, or semantic content, to the computer.

fits that are hard or impossible to predict in advance. Decentralization requires compromises: the Web had to throw away the ideal of total consistency of all of its interconnections, ushering in the infamous message “Error 404: Not Found” but allowing unchecked exponential growth.

Knowledge Representation

FOR THE SEMANTIC WEB to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning. Artificial-intelligence researchers have studied such sys-

tems since long before the Web was developed. Knowledge representation, as this technology is often called, is currently in a state comparable to that of hypertext before the advent of the Web: it is clearly a good idea, and some very nice demonstrations exist, but it has not yet changed the world. It contains the seeds of important applications, but to realize its full potential it must be linked into a single global system.

Traditional knowledge-representation systems typically have been centralized, requiring everyone to share exactly the same definition of common concepts

Glossary

HTML: Hypertext Markup Language. The language used to encode formatting, links and other features on Web pages. Uses standardized “tags” such as <H1> and <BODY> whose meaning and interpretation is set universally by the World Wide Web Consortium.

XML: eXtensible Markup Language. A markup language like HTML that lets individuals define and use their own tags. XML has no built-in mechanism to convey the meaning of the user’s new tags to other users.

RESOURCE: Web jargon for any entity. Includes Web pages, parts of a Web page, devices, people and more.

URL: Uniform Resource Locator. The familiar codes (such as <http://www.sciam.com/index.html>) that are used in hyperlinks.

URI: Universal Resource Identifier. URIs are the most familiar type of URI. A URI defines or specifies an entity, not necessarily by naming its location on the Web.

RDF: Resource Description Framework. A scheme for defining information on the Web. RDF provides the technology for expressing the meaning of terms and concepts in a form that computers can readily process. RDF can use XML for its syntax and URIs to specify entities, concepts, properties and relations.

ONTOLOGIES: Collections of statements written in a language such as RDF that define the relations between concepts and specify logical rules for reasoning about them. Computers will “understand” the meaning of semantic data on a Web page by following links to specified ontologies.

AGENT: A piece of software that runs without direct human control or constant supervision to accomplish goals provided by a user. Agents typically collect, filter and process information found on the Web, sometimes with the help of other agents.

SERVICE DISCOVERY: The process of locating an agent or automated Web-based service that will perform a required function. Semantics will enable agents to describe to one another precisely what function they carry out and what input data are needed.

such as “parent” or “vehicle.” But central control is stifling, and increasing the size and scope of such a system rapidly becomes unmanageable.

Moreover, these systems usually carefully limit the questions that can be asked so that the computer can answer reliably—or answer at all. The problem is reminiscent of Gödel’s theorem from mathematics: any system that is complex enough to be useful also encompasses unanswerable questions, much like sophisticated versions of the basic paradox “This sentence is false.” To avoid such problems, traditional knowledge-representation systems generally each had their own narrow and idiosyncratic set of rules for making inferences about their data. For example, a genealogy system, acting on a database of family trees, might include the rule “a wife of an uncle is an aunt.” Even if the data could be transferred from one system to another, the rules, existing in a completely different form, usually could not.

Semantic Web researchers, in contrast,

accept that paradoxes and unanswerable questions are a price that must be paid to achieve versatility. We make the language for the rules as expressive as needed to allow the Web to reason as widely as desired. This philosophy is similar to that of the conventional Web: early in the Web’s development, detractors pointed out that it could never be a well-organized library; without a central database and tree structure, one would never be sure of finding everything. They were right. But the expressive power of the system made vast amounts of information available, and search engines (which would have seemed quite impractical a decade ago) now pro-

duce remarkably complete indices of a lot of the material out there.

The challenge of the Semantic Web, therefore, is to provide a language that expresses both data and rules for reasoning about the data and that allows rules from any existing knowledge-representation system to be exported onto the Web.

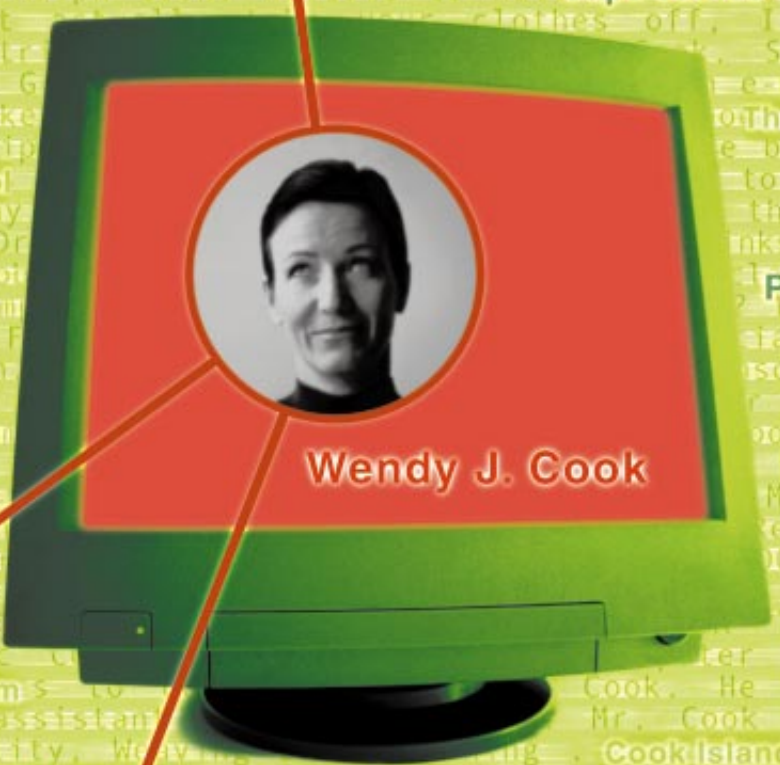
Adding logic to the Web—the means to use rules to make inferences, choose courses of action and answer questions—is the task before the Semantic Web community at the moment. A mixture of mathematical and engineering decisions complicate this task. The logic must be powerful enough to describe complex properties of objects but not so powerful that agents can be tricked by being asked to consider a paradox. Fortunately, a large majority of the information we want to express is along the lines of “a hex-head bolt is a type of machine bolt,” which is readily written in existing languages with a little extra vocabulary.

Two important technologies for developing the Semantic Web are already in place: eXtensible Markup Language (XML) and the Resource Description Framework (RDF). XML lets everyone create their own tags—hidden labels such as <zip code> or <alma mater> that annotate Web pages or sections of text on a page. Scripts, or programs, can make use of these tags in sophisticated ways, but the script writer has to know what the page writer uses each tag for. In short, XML allows users to add arbitrary structure to their documents but says nothing about what the structures mean [see “XML and the Second-Generation Web,” by Jon Bosak and Tim Bray; *SCIENTIFIC AMERICAN*, May 1999].

Meaning is expressed by RDF, which encodes it in sets of triples, each triple being rather like the subject, verb and object of an elementary sentence. These triples

ELABORATE, PRECISE AUTOMATED SEARCHES will be possible when semantics are widespread on the Web. Here a search program correctly locates a person based on an assortment of partially remembered knowledge: her last name is “Cook,” she works for a company on your client list, and she has a son attending your alma mater, Avondale University. The correct combination of that information does not reside on a single Web page, but semantics make it easier for a program to discern the elements on various pages, understand relations such as “Mike Cook is a child of Wendy Cook” and piece them together reliably. More generally, semantics will enable complicated processes and transactions to be carried out automatically.

Mr. Cook... You probably don't remember me, you were a class behind me in Russ
Cook's class prompt you to venture into all of this? Wonder where he is
Cook 1", Exhibits, Figure 1, ECG, Contact Monica.A.Garlin@uth.tmc.edu for qu
ing this information, ... sponsored by: Office ... sp
g with Mr. Cook. Good ol' boy network crucial for consulting busines
he Subject: consomme At 14:39:58 ... network
Mr. Cook's Class - Click for Grades, You MUST enter the st
EXACTLY as in the teacher's manual
Craig Cook, ... Thanks, Mr. Cook, A special
Home Page | ... Mr. Cook, contact
Livro de Receitas Eletrônica ... Aitutaki - The Blue Lagoon
Livre de qu'Phineas Cook Picture Framing ... Member ... Mr. Tim Cook, contact
Cherwell Scientific Com ... Enter Apps page, A stu
er App's voluntar ... for Mr. Cook. He is the athletic
school and ... Cook Inlet Tribal Council's Superior cotton w
ance excellent quality. Weaving and Finishing ... Mike Cook's We
arson High Main Page, Departments Page, Send to Mr. Cook. This page is
by Jessica Posey, Copyright 2000, Mike Cook, ... De Mr Cook em 06. julh
226: Para quem sabe cozinhar, Ingredientes: 1 peru de uns 5kg, pa
s 1 Garrafa de whisky ... a gosto ... Alursula Cook Fan Page's Mowing th
pal's Message, Mr. Ralph Cook, Welcome to Wal ... On behalf of
OK: So they didn't ... clothes off, I'm speaking of phy
sing you? ... Distr ... School: Gulf Stream
on: Gulf Stream FL, G ... e-mail to Mr. Bryan Co
Cook Ursula Cook ... The Inquisitive Cook: I
Cook Books Grammar tip ... e built and maintained
o much, Wendy J. Cool ... to say to Mr Mugabe Z
il report Friday Hay ... the government has m
chive, Mr Cook and Dr ... nks, Mr. Cook Special S
M'Rarotonga, You prof ... 2 Wonder where he is
Cook Phineas Cook ... taugh, tmc.edu for
Cook Strait Ferry ... sored by: AIG Insuran
ing this information ... Cor Cook Forest stit
g with Mr. Cook, ... ook Our cook@mailexc
he Subject: consomme ... M'Bedazzled - the st
Mr. Cook's ... Mr. Cook, A special
EXACTLY as in the ... return to the: Dec
Craig Cook, ... 2000, Media: Compar
Home Page | ... stas próprias re
Livro de Receitas ... Tim Cook, contact
Livre de qu' ... er Apps page, A stu
Cherwell Scientific ... Cook, He is the athletic
er A Hunterian Museum's ... Mr. Cook Superior cotton v
e school and ... Cook Islands Cook County's We
ance excellent quality. Weaving ... Cook university: Page, Send to Mr. Cook. This page is
by Jessica Posey, Copyright 2000, Mike Cook, ... em 06. julh
226: Para quem sabe cozinhar, Ingredientes: 1 peru de uns 5kg, pa
s 1 Garrafa de whisky ... a gosto ... Peter Cook and Bradley Moore Cook Mowing th
pal's Message, Mr. Ralph Cook, Welcome to Wal ... of phy
OK: So Cook Bros Light Machinery take ... for Mr. Bryan Cook, school: Gulf Stream
sing you? ... s for Mr. Bryan Cook, school: Gulf Stream
on: Gulf Stream FL, ... Fudge, Cook & Torte Chartered account as mail to Mr. Bryan Co
Cook is the caretaker, being the caretaker of ... is a ... I
ok's last ... ngabe Zimbabwe
Friday May 5, 2000 ... the government has ma
chive, Mr Cook and ... r Stran le School are cho
Mr. Cook, ... you here a class behind me in Russ
Cook's class prompt you to venture into all of this? Wonder where he is
Cook 1", Exhibits, Figure 1, ECG, Contact Monica.A.Garlin@uth.tmc.edu for qu
ing this information, ... sponsored by: Office ... sp
g with Mr. Cook. Good ol' boy network crucial for consulting busines
he Subject: consomme At 14:39:58 ... network
Mr. Cook's Class - Click for Grades, You MUST enter the st
EXACTLY as in the teacher's manual



is client of

works for

lives in

is parent of

is student at

Me

Fryer Mechanicals

Yourish Cybernetics

desJardins A.I. Co.

McGarry Automata Inc.

Bob Stacy Robotics

Stauffer Androids Ltd

Capt. James Cook

Ursula Cook

Peter Cook

Wendy J. Cook

Phineas Cook

Johannesburg

Avondale University

Greg Cook

Kangethe Gikonyo

Doranna Hatsuda

Fiona Cook

Mike Cook

Hitesh Patel

Ben Wright

Matiu Potiki

Jane Chang

Simon Brown

can be written using XML tags. In RDF, a document makes assertions that particular things (people, Web pages or whatever) have properties (such as “is a sister of,” “is the author of”) with certain values (another person, another Web page). This structure turns out to be a natural way to describe the vast majority of the data processed by machines. Subject and object are each identified by a Universal Resource Identifier (URI), just as used in

guished from an address that is a speech.

The triples of RDF form webs of information about related things. Because RDF uses URIs to encode this information in a document, the URIs ensure that concepts are not just words in a document but are tied to a unique definition that everyone can find on the Web. For example, imagine that we have access to a variety of databases with information about people, including their addresses.

ontology is a theory about the nature of existence, of what types of things exist; ontology as a discipline studies such theories. Artificial-intelligence and Web researchers have co-opted the term for their own jargon, and for them an ontology is a document or file that formally defines the relations among terms. The most typical kind of ontology for the Web has a taxonomy and a set of inference rules.

The taxonomy defines classes of ob-

The Semantic Web will enable machines to COMPREHEND semantic documents and data, not human speech and writings.

a link on a Web page. (URLs, Uniform Resource Locators, are the most common type of URI.) The verbs are also identified by URIs, which enables anyone to define a new concept, a new verb, just by defining a URI for it somewhere on the Web.

Human language thrives when using the same term to mean somewhat different things, but automation does not. Imagine that I hire a clown messenger service to deliver balloons to my customers on their birthdays. Unfortunately, the service transfers the addresses from my database to its database, not knowing that the “addresses” in mine are where bills are sent and that many of them are post office boxes. My hired clowns end up entertaining a number of postal workers—not necessarily a bad thing but certainly not the intended effect. Using a different URI for each specific concept solves that problem. An address that is a mailing address can be distinguished from one that is a street address, and both can be distin-

If we want to find people living in a specific zip code, we need to know which fields in each database represent names and which represent zip codes. RDF can specify that “(field 5 in database A) (is a field of type) (zip code),” using URIs rather than phrases for each term.

Ontologies

OF COURSE, THIS IS NOT the end of the story, because two databases may use different identifiers for what is in fact the same concept, such as **zip code**. A program that wants to compare or combine information across the two databases has to know that these two terms are being used to mean the same thing. Ideally, the program must have a way to discover such common meanings for whatever databases it encounters.

A solution to this problem is provided by the third basic component of the Semantic Web, collections of information called ontologies. In philosophy, an

ontology is a theory about the nature of existence, of what types of things exist; ontology as a discipline studies such theories. Artificial-intelligence and Web researchers have co-opted the term for their own jargon, and for them an ontology is a document or file that formally defines the relations among terms. The most typical kind of ontology for the Web has a taxonomy and a set of inference rules. The taxonomy defines classes of ob-

jects and relations among them. For example, an **address** may be defined as a type of **location**, and **city codes** may be defined to apply only to **locations**, and so on. Classes, subclasses and relations among entities are a very powerful tool for Web use. We can express a large number of relations among entities by assigning properties to classes and allowing subclasses to inherit such properties. If **city codes** must be of type **city** and cities generally have Web sites, we can discuss the Web site associated with a **city code** even if no database links a city code directly to a Web site.

Inference rules in ontologies supply further power. An ontology may express the rule “If a city code is associated with a state code, and an address uses that city code, then that address has the associated state code.” A program could then readily deduce, for instance, that a Cornell University address, being in Ithaca, must be in New York State, which is in the U.S., and therefore should be formatted to U.S. standards. The computer doesn’t truly “understand” any of this information, but it can now manipulate the terms much more effectively in ways that are useful and meaningful to the human user.

With ontology pages on the Web, solutions to terminology (and other) problems begin to emerge. The meaning of terms or XML codes used on a Web page can be defined by pointers from the page to an ontology. Of course, the same problems as before now arise if I point to an

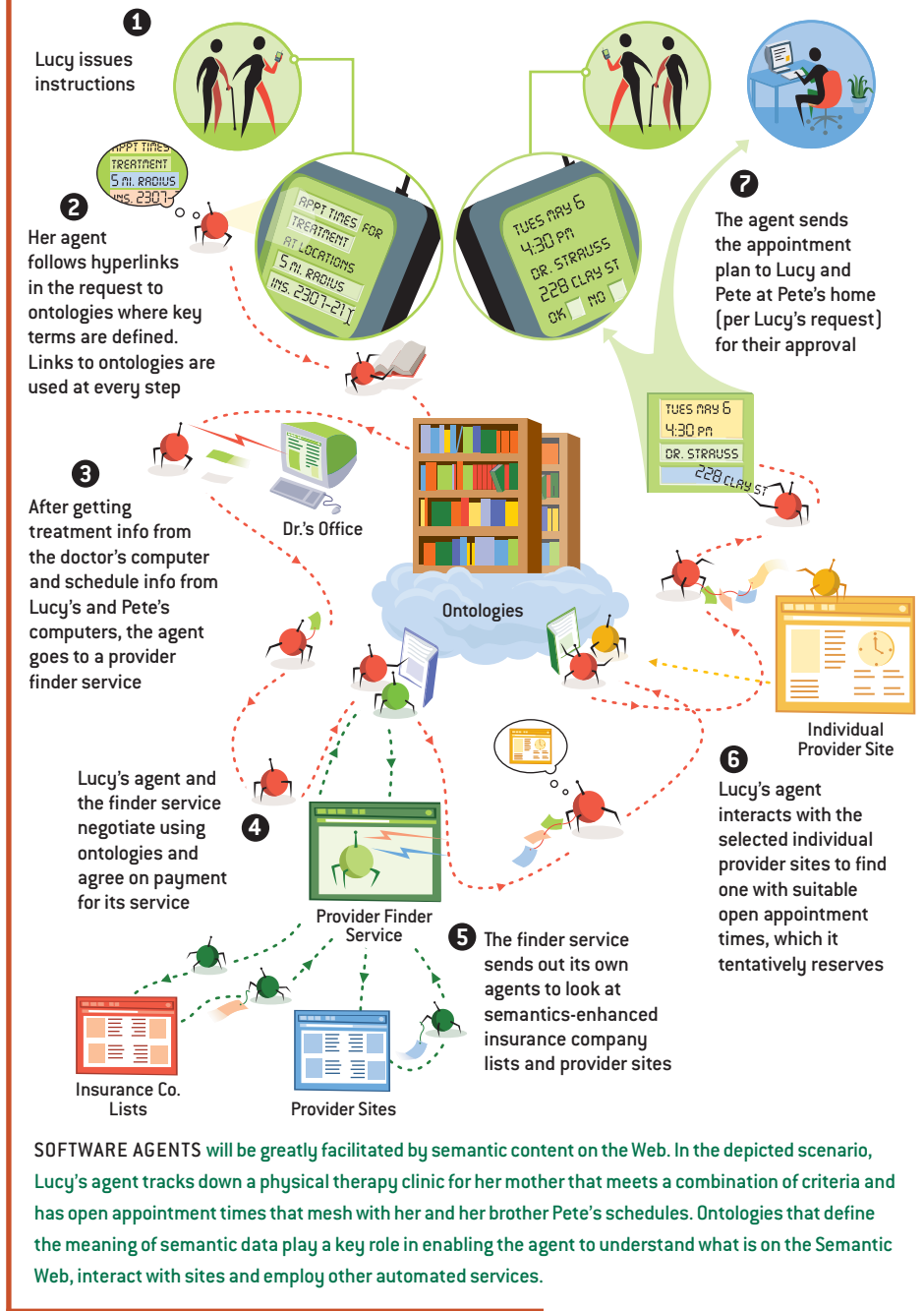
TIM BERNERS-LEE, JAMES HENDLER and ORA LASSILA are individually and collectively obsessed with the potential of Semantic Web technology. Berners-Lee is director of the World Wide Web Consortium (W3C) and a researcher at the Laboratory for Computer Science at the Massachusetts Institute of Technology. When he invented the Web in 1989, he intended it to carry more semantics than became common practice. Hendler is professor of computer science at the University of Maryland at College Park, where he has been doing research on knowledge representation in a Web context for a number of years. He and his graduate research group developed SHOE, the first Web-based knowledge representation language to demonstrate many of the agent capabilities described in this article. Hendler is also responsible for agent-based computing research at the Defense Advanced Research Projects Agency (DARPA) in Arlington, Va. Lassila is a research fellow at the Nokia Research Center in Boston, chief scientist of Nokia Venture Partners and a member of the W3C Advisory Board. Frustrated with the difficulty of building agents and automating tasks on the Web, he co-authored W3C’s RDF specification, which serves as the foundation for many current Semantic Web efforts.

ontology that defines **addresses** as containing a **zip code** and you point to one that uses **postal code**. This kind of confusion can be resolved if ontologies (or other Web services) provide equivalence relations: one or both of our ontologies may contain the information that my **zip code** is equivalent to your **postal code**.

Our scheme for sending in the clowns to entertain my customers is partially solved when the two databases point to different definitions of **address**. The program, using distinct URIs for different concepts of **address**, will not confuse them and in fact will need to discover that the concepts are related at all. The program could then use a service that takes a list of postal **addresses** (defined in the first ontology) and converts it into a list of physical **addresses** (the second ontology) by recognizing and removing post office boxes and other unsuitable addresses. The structure and semantics provided by ontologies make it easier for an entrepreneur to provide such a service and can make its use completely transparent.

Ontologies can enhance the functioning of the Web in many ways. They can be used in a simple fashion to improve the accuracy of Web searches—the search program can look for only those pages that refer to a precise concept instead of all the ones using ambiguous keywords. More advanced applications will use ontologies to relate the information on a page to the associated knowledge structures and inference rules. An example of a page marked up for such use is online at <http://www.cs.umd.edu/~hendler>. If you send your Web browser to that page, you will see the normal Web page entitled “Dr. James A. Hendler.” As a human, you can readily find the link to a short biographical note and read there that Hendler received his Ph.D. from Brown University. A computer program trying to find such information, however, would have to be very complex to guess that this information might be in a biography and to understand the English language used there.

For computers, the page is linked to an ontology page that defines information about computer science depart-



SOFTWARE AGENTS will be greatly facilitated by semantic content on the Web. In the depicted scenario, Lucy's agent tracks down a physical therapy clinic for her mother that meets a combination of criteria and has open appointment times that mesh with her and her brother Pete's schedules. Ontologies that define the meaning of semantic data play a key role in enabling the agent to understand what is on the Semantic Web, interact with sites and employ other automated services.

ments. For instance, professors work at universities and they generally have doctorates. Further markup on the page (not displayed by the typical Web browser) uses the ontology's concepts to specify that Hendler received his Ph.D. from the entity described at the URI <http://www.brown.edu/>—the Web page for Brown. Computers can also find that Hendler is a member of a particular research project, has a particular e-mail address, and so on. All that information is readily processed by a computer and could be used to answer queries (such as where Dr. Hendler received his degree) that cur-

rently would require a human to sift through the content of various pages turned up by a search engine.

In addition, this markup makes it much easier to develop programs that can tackle complicated questions whose answers do not reside on a single Web page. Suppose you wish to find the Ms. Cook you met at a trade conference last year. You don't remember her first name, but you remember that she worked for one of your clients and that her son was a student at your alma mater. An intelligent search program can sift through all the pages of people whose name is

What Is the Killer App?



AFTER WE GIVE a presentation about the Semantic Web, we're often asked, "Okay, so what is the killer application of the Semantic Web?" The "killer app" of any technology, of course, is the application that brings a user to investigate the technology and start using it. The transistor radio was a killer app of transistors, and the cell phone is a killer app of wireless technology.

So what do we answer? "The Semantic Web *is* the killer app."

At this point we're likely to be told we're crazy, so we ask a question in turn: "Well, what's the killer app of the World Wide Web?" Now we're being stared at kind of fish-eyed, so we answer ourselves: "The Web is the killer app of the Internet. The Semantic Web is another killer app of that magnitude."

The point here is that the abilities of the Semantic Web are too general to be thought about in terms of solving one key problem or creating one essential gizmo. It will have uses we haven't dreamed of.

Nevertheless, we can foresee some disarming (if not actually killer) apps that will drive initial use. Online catalogs with semantic markup will benefit both buyers and sellers. Electronic commerce transactions will be easier for small businesses to set up securely with greater autonomy. And one final example: you make reservations for an extended trip abroad. The airlines, hotels, soccer stadiums and so on return confirmations with semantic markup. All the schedules load directly into your date book and all the expenses directly into your accounting program, no matter what semantics-enabled software you use. No more laborious cutting and pasting from e-mail. No need for all the businesses to supply the data in half a dozen different formats or to create and impose their own standard format.

"Cook" (sidestepping all the pages relating to cooks, cooking, the Cook Islands and so forth), find the ones that mention working for a company that's on your list of clients and follow links to Web pages of their children to track down if any are in school at the right place.

Agents

THE REAL POWER of the Semantic Web will be realized when people create many programs that collect Web content from diverse sources, process the information and exchange the results with other programs. The effectiveness of such software agents will increase exponentially as more machine-readable Web content and automated services (including other agents) become available. The Semantic Web promotes this synergy: even agents that were not expressly designed to work together can transfer data among themselves when the data come with semantics.

An important facet of agents' functioning will be the exchange of "proofs"

written in the Semantic Web's unifying language (the language that expresses logical inferences made using rules and information such as those specified by ontologies). For example, suppose Ms. Cook's contact information has been located by an online service, and to your great surprise it places her in Johannesburg. Naturally, you want to check this, so your computer asks the service for a proof of its answer, which it promptly provides by translating its internal reasoning into the Semantic Web's unifying language. An inference engine in your computer readily verifies that this Ms. Cook indeed matches the one you were seeking, and it can show you the relevant Web pages if you still have doubts. Although they are still far from plumbing the depths of the Semantic Web's potential, some programs can already exchange proofs in this way, using the current preliminary versions of the unifying language.

Another vital feature will be digital signatures, which are encrypted blocks of

data that computers and agents can use to verify that the attached information has been provided by a specific trusted source. You want to be quite sure that a statement sent to your accounting program that you owe money to an online retailer is not a forgery generated by the computer-savvy teenager next door. Agents should be skeptical of assertions that they read on the Semantic Web until they have checked the sources of information. (We wish more *people* would learn to do this on the Web as it is!)

Many automated Web-based services already exist without semantics, but other programs such as agents have no way to locate one that will perform a specific function. This process, called service discovery, can happen only when there is a common language to describe a service in a way that lets other agents "understand" both the function offered and how to take advantage of it. Services and agents can advertise their function by, for example, depositing such descriptions in directories analogous to the Yellow Pages.

Some low-level service-discovery schemes are currently available, such as Microsoft's Universal Plug and Play, which focuses on connecting different types of devices, and Sun Microsystems's Jini, which aims to connect services. These initiatives, however, attack the problem at a structural or syntactic level and rely heavily on standardization of a predetermined set of functionality descriptions. Standardization can only go so far, because we can't anticipate all possible future needs.

The Semantic Web, in contrast, is more flexible. The consumer and producer agents can reach a shared understanding by exchanging ontologies, which provide the vocabulary needed for discussion. Agents can even "bootstrap" new reasoning capabilities when they discover new ontologies. Semantics also makes it easier to take advantage of a service that only partially matches a request.

A typical process will involve the creation of a "value chain" in which sub-assemblies of information are passed from one agent to another, each one "adding value," to construct the final product requested by the end user. Make no mistake:

to create complicated value chains automatically on demand, some agents will exploit artificial-intelligence technologies in addition to the Semantic Web. But the Semantic Web will provide the foundations and the framework to make such technologies more feasible.

Putting all these features together results in the abilities exhibited by Pete's and Lucy's agents in the scenario that opened this article. Their agents would

Pete answers his phone and the stereo sound is turned down. Instead of having to program each specific appliance, he could program such a function once and for all to cover every *local* device that advertises having a *volume control*—the TV, the DVD player and even the media players on the laptop that he brought home from work this one evening.

The first concrete steps have already been taken in this area, with work on de-

Human endeavor is caught in an eternal tension between the effectiveness of small groups acting independently and the need to mesh with the wider community. A small group can innovate rapidly and efficiently, but this produces a subculture whose concepts are not understood by others. Coordinating actions across a large group, however, is painfully slow and takes an enormous amount of communication. The world works

Properly designed, the Semantic Web can assist the **E V O L U T I O N** of human knowledge as a whole.

have delegated the task in piecemeal fashion to other services and agents discovered through service advertisements. For example, they could have used a *trusted* service to take a list of *providers* and determine which of them are *in-plan* for a specified *insurance plan* and *course of treatment*. The list of providers would have been supplied by another search service, et cetera. These activities formed chains in which a large amount of data distributed across the Web (and almost worthless in that form) was progressively reduced to the small amount of data of high value to Pete and Lucy—a plan of appointments to fit their schedules and other requirements.

In the next step, the Semantic Web will break out of the virtual realm and extend into our physical world. URIs can point to anything, including physical entities, which means we can use the RDF language to describe devices such as cell phones and TVs. Such devices can advertise their functionality—what they can do and how they are controlled—much like software agents. Being much more flexible than low-level schemes such as Universal Plug and Play, such a semantic approach opens up a world of exciting possibilities.

For instance, what today is called home automation requires careful configuration for appliances to work together. Semantic descriptions of device capabilities and functionality will let us achieve such automation with minimal human intervention. A trivial example occurs when

developing a standard for describing functional capabilities of devices (such as screen sizes) and user preferences. Built on RDF, this standard is called Composite Capability/Preference Profile (CC/PP). Initially it will let cell phones and other nonstandard Web clients describe their characteristics so that Web content can be tailored for them on the fly. Later, when we add the full versatility of languages for handling ontologies and logic, devices could automatically seek out and employ services and other devices for added information or functionality. It is not hard to imagine your Web-enabled microwave oven consulting the frozen-food manufacturer's Web site for optimal cooking parameters.

Evolution of Knowledge

THE SEMANTIC WEB is not “merely” the tool for conducting individual tasks that we have discussed so far. In addition, if properly designed, the Semantic Web can assist the evolution of human knowledge as a whole.

across the spectrum between these extremes, with a tendency to start small—from the personal idea—and move toward a wider understanding over time.

An essential process is the joining together of subcultures when a wider common language is needed. Often two groups independently develop very similar concepts, and describing the relation between them brings great benefits. Like a Finnish-English dictionary, or a weights-and-measures conversion table, the relations allow communication and collaboration even when the commonality of concept has not (yet) led to a commonality of terms.

The Semantic Web, in naming every concept simply by a URI, lets anyone express new concepts that they invent with minimal effort. Its unifying logical language will enable these concepts to be progressively linked into a universal Web. This structure will open up the knowledge and workings of humankind to meaningful analysis by software agents, providing a new class of tools by which we can live, work and learn together. **SA**

MORE TO EXPLORE

Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor. Tim Berners-Lee, with Mark Fischetti. Harper San Francisco, 1999.

An enhanced version of this article is on the Scientific American Web site, with additional material and links.

World Wide Web Consortium (W3C): www.w3.org/

W3C Semantic Web Activity: www.w3.org/2001/sw/

An introduction to ontologies: www.SemanticWeb.org/knowmarkup.html

Simple HTML Ontology Extensions Frequently Asked Questions (SHOE FAQ): www.cs.umd.edu/projects/plus/SHOE/faq.html

DARPA Agent Markup Language (DAML) home page: www.daml.org/