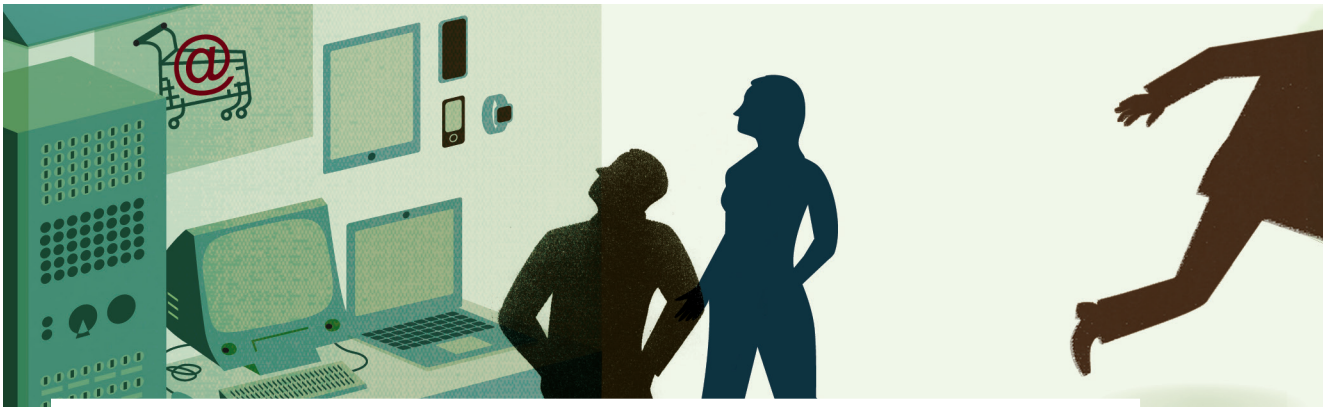

The Web That Extends beyond the World

Vishnu S. Pendyala, Santa Clara University
Simon S.Y. Shim, San Jose State University
Christoph Bussler, Oracle

The Web evolved much more rapidly than was probably envisioned when it was invented two decades ago. The authors present their vision for the Web's next two decades by identifying important research directions that can help extend its reach further.

"The Internet will disappear."
—Eric Schmidt, executive chairman of Google

Although it didn't exist 25 years ago, the World Wide Web is now an indispensable part of life for many. But in the near future, as Google's executive chairman Eric Schmidt predicted at the World Economic Forum in January 2015, the Web and its applications will be so well integrated into our daily lives that we will no longer notice their presence. The same happened with electronic devices. From toasters and alarm clocks to the cars we drive, electronic chips control behavior all around us; we are so used to these devices that they are now almost invisible to our consciousness.

The Web has even greater potential than electronic devices to impact the world, due to its ability to connect and expose those devices' services to bring about great synergy. With the advent of wearables and the Internet of Things—and their applications—the Web is set to permeate every nook and cranny of the world. Web interactions will no longer be limited to the browser—users will be able to speak into devices and hear back from them seamlessly. The Web will also significantly extend its outreach into space.

In this article, we examine some of the enabling factors, trends, and requirements that will make these predictions come true. These new features might need protocol, application, or infrastructure changes—considerations we leave for future research. The Web's underlying infrastructure shapes its applications, and these applications

0018-9162/15/$31.00 © 2015 IEEE

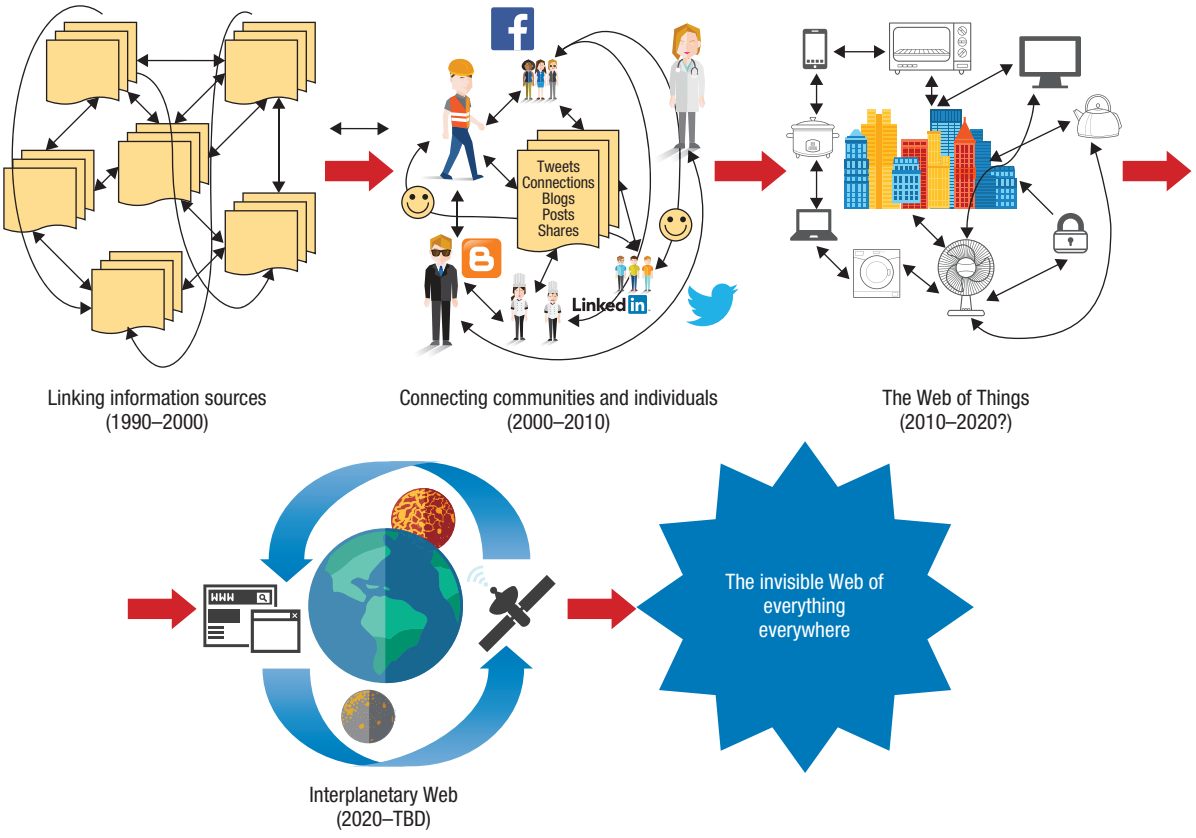

FIGURE 1. Evolution of the Web. The Web is evolving from an interconnection of static information sources to a programmable interconnection of static and dynamic information, people, communities, things, and—eventually—planets.

give the Web its character. Thus, we take a conflated view of the Web when discussing its future.

THE EVOLVING WEB

The Web changes rapidly and has the potential to serve as a conduit for tremendous computing power from almost anywhere. We discussed this vision of the Web as a ubiquitous computer in the September 2009 issue of this magazine.[1] As depicted in Figure 1, the Web is evolving from an interconnection of static information sources to a programmable interconnection of static and dynamic information, people, communities, things, and—eventually—planets. This evolution is a testament to the amount of research focus the Web has received over the past two decades and will continue to receive.

There is still plenty of scope for evolving the Web. Knowledge representation is key to building intelligent Web applications. As the need for machine-to-machine communication grows with the increasing ubiquity of the WoT, machine-readable representations such as XML, JSON, and the Semantic Web are heading in the right direction, but there is substantially more reasoning necessary to align the Web more closely with the real world. The real world is characterized by uncertainty, but so far, the Web does not reflect this fact.

Reasoning under uncertain circumstances is an important research goal. The current knowledge representations available to Web protocols are incapable of meeting this requirement. The underlying assumption of the processing done on the Web is that the information available to it is all true. In "Streaming the Web: Reasoning over Dynamic Data," Alessandro Margara and his colleagues discuss the importance of handling uncertainty in streaming data on the Web and conclude that current systems for doing so are inadequate.[2] Dealing with uncertainty in the information represented on the Web will be a notable milestone in its evolution.

Multiple zettabytes of data on the Web are locked up in suboptimal structures. An important project would be to construct a knowledge base from that data; an appropriate representation of such knowledge could facilitate the creation of Web applications that can reason and draw conclusions. Advances in text mining and knowledge representation could make this possible. Knowledge representation has been a formidable challenge for computer scientists, owing to what is currently possible versus what is potentially possible. Significant innovation

AUGMENTED REALITY

Augmented reality (AR) will be a major enabler in the Web's evolution. Its applications promise to take quality of life to the next level. Thus, it is no surprise that Microsoft is projecting HoloLens as the next wave in personal computing (www.microsoft.com/microsoft-hololens). Google Glass popularized AR and continues to do so. AR technology augments the real world by projecting virtual information in a user's vicinity. The information is mostly visual at this time but can include smell, touch, and hearing. HoloLens augments reality by creating holograms, combining the real and virtual worlds.

In conjunction with paradigms like the WoT and the interplanetary Internet, Web-based AR applications could offer a very different way of experiencing the universe. Users could get a near-real-life visualization of the structures in outer space and on other planets. Cosmic activity could be replayed through AR. Prototypes could be built and experimented with to gain better insights or even control cosmic entities. For more lasting prototypes, a virtual object could even become real using a 3D printer. Objects on Mars could be compared with similar ones on Earth.

Certain ethical and legal considerations could impede AR adoption, particularly regarding privacy. AR could be an impediment to safe driving. It could be a major distraction, obsession, and even addiction. As Franziska Roesner and her colleagues discuss in "Augmented Reality: Hard Problems of Law and Policy," AR could lead to illegal discrimination, violation of intellectual property rights, and misleading information (*Proc. ACM Int'l Joint Conf. Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, pp. 1283–1288). The potential for misuse is high. However, once set in motion, technology never stops. The valid concerns will be addressed in time, and AR will likely become the most disruptive of the user-experience technologies.

in this area will be disruptive, particularly regarding the vast amounts of data that will be automatically generated from the innumerable sensors embedded in the Web of Things (WoT).

Developments in augmented reality (see the "Augmented Reality" sidebar), machine learning, and natural language processing will transform the Web into a medium that seamlessly permeates into our social fabric. To become ubiquitous, the Web should speak the language of the user and be able to communicate in ways that are best suited to the user. Speech recognition logic is already built into the Chrome browser—it should be possible in the near future for the Web I/O to happen in audio and video, in the user's native language, without having to use a keyboard.

User experience will continue to drive the Web's evolution. Web protocols might need to change to accommodate the various IO channels—such as speech—which could obviate the need for users to type URLs. Machines might continue to use URLs or URIs to access Web resources, but the internal mechanism should be abstracted away from users. The way to access information needs to evolve, as does the way in which information is harnessed.

An emerging breed of applications aggregates, summarizes, and annotates information on the Web. This category of machine-generated information could result in entire multimedia magazines produced without any human involvement. Summly—an app that delivers short summaries of news stories and was recently acquired by Yahoo—and Google's translation tool prove that statistical techniques can perform better than semantic techniques when it comes to text processing. With the wealth of statistical techniques at our disposal, it is very possible that machine-generated information on the Web could be much better than information written by humans. Big data analytics, which uses statistical toolkits to process information, holds huge promise here.

Another important aspect of the intelligent Web is its ability to learn. Current Web applications like search engines already learn from user behavior: search queries are predicted based on past queries, and advertisements are displayed based on usage. But this is minimal compared to the learning that happens in the real world. Web applications should be able to quickly learn patterns from the knowledge made available to them and use these patterns to predict the future. For example, Google X created a network of 16,000 computers that learned to recognize videos of cats on YouTube (www.wired.com/2012/06/google-x-neural-network). This ability could detect and prevent malicious Web activity or help with everyday situations; for instance, a car's computer could predict when a freeway is clogged and reroute the driver.

THE WEB OF THINGS

The number of devices connected on the Internet already exceeds the

world's population. Cisco predicts that there will be 50 billion things connected on the Internet by 2020 (http://share.cisco.com/internet-of-things.html). These things include not only computers and smartphones but also cars, washing machines, and even cattle (www.theguardian.com/local-government-network/2011/aug/18/internet-of-things-local-government). The synergy brought about by this interconnection of things is possible only when the things' services are made accessible via Web protocols. The Web is the behind-the-scenes medium that helps harness these interconnected things to provide a rich set of applications that improve the quality of life. Once interconnected on the Web, resources including personal gadgets can be shared remotely and used more efficiently.[3] The WoT, therefore, is a significant enabler of the "disappearing" Internet.

Web-accessible, wearable smart devices are gaining popularity. Google Glass started the trend, which is expanding to smart watches, headbands that sense brain waves (www.choosemuse.com), and smart clothing, such as a shirt that detects the wearer's heart rate, speed, breathing patterns, and GPS location (www.digitaltrends.com/wearables/smart-clothing-garments-at-ces-2015-and-beyond).

Currently, several ontology schemes represent the data generated from interconnected things' sensors. These ontologies are predominantly based on the Resource Description Framework and help elucidate conclusions from semantic representations using first-order logic or machine-learning techniques. A typical conclusion is the detection of an anomaly such as an intrusion, fraud, or—in the case of a connected smart grid—a sudden change in the power supply frequency. The WoT provides the infrastructure to build applications that can arrive at such conclusions, leading to a dramatic increase in applications.

THE APPLICATION EXPLOSION

Many applications that are based on the WoT—ranging from tracking cargo to building smart cities—already exist. Efficient energy use, leak detection, intelligent buildings and parking systems, air quality monitoring, and the monitoring of buildings' structural health are all possible today with the help of the WoT. Coupled with augmented reality, the possibilities are endless: virtual universities, automated courtrooms, and point-of-care hospitals can become a reality. Analyzing the Web's big data will enable a number of applications, some of which Emilio Ferrara and his colleagues discuss in "Web Data Extraction, Applications and Techniques: A Survey."[4] This article is a good treatise on the various techniques in using Web data to build applications.

University of Southern California researchers are working on embedding microchips inside the human brain to aid cognitive functions such as memory. With such recent advances in brain–computer interfaces (BCIs), it's plausible that human and machine knowledge repositories will converge toward a common format of representation. Humans and machines can then exchange knowledge, and isolated chunks of knowledge can be connected and reasoned with on the Web.

In 2014, scientists successfully demonstrated direct communication between two human brains (www.washington.edu/news/2014/11/05/uw-study-shows-direct-brain-interface-between-humans). If brains can communicate through embedded microchips, they can potentially become yet another part of the WoT. The human brain can have an IP address and communicate directly with other human brains and machines, resulting in limitless human–machine synergy. The Web can help unleash this unbounded potential. Language barriers will be broken, and the world will become much more personalized and honest, as it will be harder to lie. Moods will be able to be predicted, psychological processes will be analyzed on the fly, and services will be provided based on this information.

Although traditionally used to help restore damaged hearing, sight, and movement, BCIs are starting to be used on healthy individuals as well, opening up a wide spectrum of applications. In "Towards Passive Brain–Computer Interfaces: Applying Brain–Computer Interface Technology to Human–Machine Systems in General," Thorsten O. Zander and Christian Kothe talk about a passive BCI, which combines BCI techniques with cognitive monitoring to understand what goes on in the human mind.[5] The information generated from such an application when connected with other Web resources can result in highly disruptive technologies: the age of the "machine-augmented mind" and the "Web-enabled brain" might not be too far away.

BCI also brings with it new problems. In "Securing the Exocortex: A Twenty-First Century Cybernetics Challenge," Tamara Bonaci and her colleagues present an information-theoretic framework to address brain spyware—BCI-enabled malicious applications that extract sensitive

information about an individual, including passwords and PINs, from neural signals.[6] Issues of security and privacy will become even more critical when BCI applications begin to run on the Web.

THE INTERPLANETARY WEB

The past decade has seen considerable research on the interplanetary Internet[7] and delay-tolerant networks in an effort to extend the Internet beyond Earth. Hardware is deployed in outer space to help us learn more about our solar system. Exposing the hardware's functionality and building applications to harness that hardware will help take the Web to the next level, transforming it into the interplanetary Internet.

Governments' monopoly on space travel is loosening, and private participation is constantly rising, indicating increased interest in and possibilities of space travel. Elon Musk of Tesla Motors and SpaceX has announced an ambitious project to place hundreds of satellites 750 miles above Earth, speeding up the transfer of data and giving better coverage to the 3 billion people who do not have Internet access. The project would also form the basis of a communications system with Mars if Musk is able to achieve his goal of establishing a human colony there.

As a natural corollary, the Web in the application layer should expand across planets, allowing for a range of interplanetary applications. With developments in robotics, augmented reality, and related fields, Web users will be able to experience different planets just by surfing the Web. Users will be able to get firsthand insight into cosmic activity, which could inspire new terrestrial inventions. By better understanding the universe, we can better control and improve our own processes.

Using the interplanetary Internet, we can work toward making other planets inhabitable by humans. Web-connected wearables will allow astronauts to communicate with their friends and family, other astronauts, and control centers elsewhere in space. If other planets become inhabitable, it is imperative that the Web extend to those planets. The Web must become the primary medium for communication and transactions among planets.

Until now, humans have had to add or delete things from the Web. But with the advent of the WoT, this should change. The WoT is based on peer-to-peer networking: smart objects in the WoT can communicate with and learn from each other. To extend the Web, the devices in the WoT should not only use the Web but also be accessible to other devices when necessary. Automatic discovery, induction, and deduction are needed to make the Web truly ubiquitous.

Due to delays in the interplanetary Internet and the possibility that some things in the Internet of Things might not be online, store and forward techniques are likely to become commonplace in communications. The future Web should conceal such mechanisms and the resulting delay to provide a seamless experience as well as preserve transactional consistency and other ACID (atomicity, consistency, isolation, durability) properties.

The WoT paradigm is likely to necessitate protocol changes that the interplanetary Internet can benefit from. For instance, currently, when a device's network is down, buffering happens locally on the device. In the future, Web protocols might need to handle buffering. This is particularly important given the unreliable communication that characterizes sensor networks and therefore the WoT. Protocols should increasingly ensure reliable packet delivery. We currently have routers that recognize XML; future Web protocols might need to recognize the data structure for better processing.

On a related note, there is a need for seamless operations on slow networks. For wider reach, the Internet needs to work in low-bandwidth locations such as airplanes and remote places in developing nations. We need to find effective ways to provide infinite bandwidth even when the network is slow. However, mechanisms such as store and forward need to be abstracted away from users.

THE HUMANITARIAN WEB

The economy continues to grow as more people join its core echelons. People are the most important economic resource. The Web will play an ever-increasing role in bringing underprivileged populations to the core echelons, possibly without these populations knowing that the Web is doing this work. Global applications—especially those with a humanitarian perspective—will have to be developed and made accessible on the Web. We already have microfinance and crowdsourcing websites, which are quite popular in developing countries. In "Improving Visibility of Humanitarian Supply Chains through Web-Based Collaboration,"[8] Mohammad Anwar Rahman discusses the concept of using the Web for supply-chain management among humanitarian organizations. There is a pressing need for more game-changing, universal Web applications for mass deployment.

Healthcare can immensely benefit from such applications. Millions

of people who do not have access to healthcare do have access to a mobile device. Web applications that can be accessed using smartphones, such as automated medical diagnosis,[9] need to be developed and deployed across the world. The Web can also help people with limited mobility or access to transportation by enabling projects like virtual universities.

THE POLITICAL WEB

Web 2.0 opened up a Pandora's box for government and politicians. Markus Schatten and his colleagues' "An Introduction to Social Semantic Web Mining & Big Data Analytics for Political Attitudes and Mentalities Research" shows how big data mining could reveal the dynamics of collective action on the Web.[10] With Web 2.0, anyone with Internet access had the potential to become a journalist by posting opinions, pictures, and videos. The Web became a more powerful medium than the press. Over the next 25 years, the omnipresent Web could become much more powerful.

In the future, government-ordered snooping will increase, and political censorship will peak (see the "Securing the Web: New Frontiers" sidebar). Web content will increasingly be bound by geographic limits. Something similar to today's virtual private networks could be implemented across the world to guard and regulate Web traffic. On the other hand, governments could facilitate the availability of data devoid of personally identifiable information for humanitarian use, such as the medical diagnosis application described in Vishnu S. Pendyala and his colleagues' "A Text Mining Approach to Automated Healthcare for the Masses."[9] It is also possible that governments could substantially invest in securing the Web or that technological policies governing the Web—such as Net neutrality—could become major differentiators in future election agendas.

THE PRIVACY PREDICAMENT

Online privacy is a common concern for many. Still, most people unknowingly give up their privacy when using the Internet. Many recent instances of privacy violation have caused substantial harm. In one of the largest hacking prosecutions, it was discovered that 160 million credit and debit card numbers were stolen. The hackers targeted 800,000 bank accounts. Companies have been caught using their customers' private data under various pretexts. Often, the data that flows through or resides on a company's infrastructure is more valuable than the infrastructure itself. The WoT and the social networking Web are making companies more powerful because of the data they generate.

Many companies help protect online privacy, but they leave much to be desired. With the government's will and technical support, it should not be too difficult to prioritize absolute privacy. Using appropriate encryption and other techniques, it should be possible to allow users to choose their audience. For instance, using Wi-Fi instead of cell towers when placing calls can make it more difficult to locate the callers, so service providers can allow users the option to choose between the two. For Wi-Fi or wired traffic, encryption should happen at the origin, and its private key should lie with the originator. However, such mechanisms severely impact the interests of the companies providing the services. It also makes it more difficult for governments to detect suspicious activity. Tradeoffs and contingencies can and should be considered.

WEB SECURITY

The security and privacy of sensitive information on the Web should be of utmost importance. Cyberattacks will become much more sophisticated and threats much more involved. Even hardware is not exempt: Trojan horses can be implanted in devices.[11] Any security breach could lead to disaster. For instance, in the case of a grid-connected nuclear power plant, a cyberattack could lead to the loss of life. The Web should be able to cope with these dynamic security challenges, particularly as it extends beyond the boundaries of Earth. If the Web is the furniture, appliances, and gadgets that we accumulate to live comfortably, Web security is the house that protects these belongings from natural and man-made disasters.

Before the Internet, corporate networks had accounts that required logins and passwords to access the network. This meant there were fewer instances of spam and malicious attacks, as the originator could be identified in seconds. There was more accountability and better control over network traffic. All this changed with the advent of the Internet. Today, the Web is dominated by anonymous and pseudonymous operations. A partial solution would be to go back to the days of stringent guidelines for gaining network access or doing anything online. Paradigm shifts like software-defined security might help intermittently, but the Web architecture might need

SECURING THE WEB: NEW FRONTIERS

The most significant change in the way we will use the Web in the future is our connectedness to a number of devices. Today, smartphones and a few other computing devices are connected to the Internet. There are massive attempts, even by governments, to gain control over these devices' data and operations. In "Highlights from Making Sense of Snowden, Part II: What's Significant in the NSA Revelations," Susan Landau talks about how the US government maneuvered security and cryptographic standards, causing a long-term adverse impact on cybersecurity worldwide (*IEEE Security & Privacy*, vol. 12, no. 1, 2014, pp. 62–64). Gary King and his colleagues' "Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation" explains the Chinese government's effort to undermine collective activity on the Web (*Science*, vol. 345, no. 6199, 2014; doi: 10.1126/science.1251722). Powerful technology gives rise to various interested parties that try to subvert that technology.

This problem will only be compounded in the future as the number of devices people rely on as part of their daily routine increases. The Web's extension to other planets and its increasing influence on the political agenda further complicate the problem. A significant number of devices that connect to the Web are likely to be programmable, powerful, and pervasive. These devices' information flow goes two ways: the entity gets to know users through the devices they use, and users get to know the entity that is using their information. The latter is often benign, although malicious users exist. The issue is mostly malicious entities gaining control over users' data and operations.

Today, even hardware can't be trusted, as it is susceptible to malicious Trojan horse viruses. Software has many more holes that need to be plugged. Encryption might seem like a silver bullet, but it is not particularly aligned with most business models or the National Security Agency's agenda. We need better solutions. The most vulnerable entity when using connected devices is one's identity. If a device is compromised, it could very easily lead to identity theft. One way to avoid this is to isolate the identity from the device. For instance, cell phones in some Asian countries are prepaid and therefore are not personally identifiable. Swiss bank accounts also are not personally identifiable—a unique ID and a password are required to gain access. Similar techniques could be used to protect users from identity theft. However, all of this comes with tradeoffs that need to be carefully considered.

to be revamped to plug loopholes and secure the Web from the ground up.

Privacy and security are of paramount importance in these scenarios (see "The Privacy Predicament" sidebar). Not all devices can automatically become accessible on the Web, even when they participate in extending it. A router is one such device. Without exception, users should be able to control who can access their information, devices, and any other resources they wish to make available online. Even governments, Internet providers, and hosting companies should not be able to gain access without an owner's authorization. This will require protocol and legal changes.

The ideas presented in this article give a broad overview of some of the research directions for the Web that should be probed more deeply. The future we predict here depends on certain inventions and key breakthroughs in other areas such as BCI. Advances in lower layers of the protocol stack, machine-to-machine interaction, and representation of machine-readable data also play a prominent role in this respect. The weakest links, however, are the security and privacy aspects of the Web as they make it vulnerable. Technology has been advancing in leaps and bounds to improve the quality of life. The power of the Web lies in connecting the services of this technology with other knowledge, computing, and networking resources to unleash possibilities that stretch into space and beyond.

REFERENCES

1. V.S. Pendyala and S.S. Shim, "The Web as the Ubiquitous Computer,"

Computer, vol. 42, no. 9, 2009, pp. 90–92.
2. A. Margara et al., "Streaming the Web: Reasoning over Dynamic Data," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 25, 2014, pp. 24–44.
3. D. Guinard, M. Fischer, and V. Trifa, "Sharing Using Social Networks in a Composable WoT," *Proc. 8th IEEE Int'l Conf. Pervasive Computing and Communications Workshops* (PERCOM Workshops 10), 2010, pp. 702–707.
4. E. Ferrara et al., "Web Data Extraction, Applications and Techniques: A Survey," *Knowledge-Based Systems*, vol. 70, 2014, pp. 301–323.
5. T.O. Zander and C. Kothe, "Towards Passive Brain–Computer Interfaces: Applying Brain–Computer Interface Technology to Human–Machine Systems in General," *J. Neural Eng.*, vol. 8, no. 2, 2011; doi:10.1088/1741-2560/8/2/025005.
6. T. Bonaci et al., "Securing the Exocortex: A Twenty-First Century Cybernetics Challenge," *IEEE Conf. Norbert Wiener in the 21st Century* (21CW 14), 2014, pp. 1–8.
7. J. Mukherjee and B. Ramamurthy, "The Interplanetary Internet Implemented on a Terrestrial Testbed," *IEEE Int'l Conf. Comm.* (ICC 13), 2013, pp. 4298–4303.
8. M.A. Rahman, "Improving Visibility of Humanitarian Supply Chains through Web-Based Collaboration," *User-Centric Technology Design for Nonprofit and Civic Engagements*, Springer, 2014, pp. 69–87.
9. V.S. Pendyala et al., "A Text Mining Approach to Automated Healthcare for the Masses," *Proc. IEEE Global Humanitarian Technology Conf.* (GHTC 14), 2014, pp. 28–35.
10. M. Schatten, J. Ševa, and B.O. Đurić, "An Introduction to Social Semantic Web Mining & Big Data Analytics for Political Attitudes and Mentalities Research," *European Q. Political Attitudes and Mentalities EQPAM*, vol. 4, no. 1, 2015, pp. 40–62.
11. E. Potkonjak et al., "Hardware Trojan Horse Detection Using Gate-Level Characterization," *Proc. 46th Ann. ACM/IEEE Design Automation Conf.* (DAC 09), 2009, pp. 688–693.

ABOUT THE AUTHORS

VISHNU S. PENDYALA is pursuing his PhD in computer engineering at Santa Clara University. His research interests include machine learning, Semantic Web, and applying technology to humanitarian causes. He is a senior member of the IEEE Computer Society. Contact him at pendyala@computer.org.

SIMON S.Y. SHIM is a professor in the Computer Engineering Department at San Jose State University. His current research interests include Internet computing, databases, machine learning, and information retrieval. Shim received a PhD in computer science from the University of Minnesota. He is a member of the IEEE Computer Society. Contact him at simon.shim@sjsu.edu.

CHRISTOPH BUSSLER is Consulting Member Technical Staff at Oracle Corporation's cloud division. His research interests include cloud technology, nonstandard database models, Web services, enterprise integration, and workflow management. Bussler received his PhD in computer science from Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany. He is a member of the IEEE Computer Society. Contact him at christoph.bussler@oracle.com.

DISCLAIMER

The views expressed in this article do not necessarily reflect the views of Oracle Corporation.