Our New Website Has Launched!

Some of the many new features include:

- **Digital Library**—Improved search helps you more easily navigate 700k+ articles in the IEEE Computer Society Digital Library.
- **Conference Calendar**—Plan your event learning experiences with new filters for 200+ conferences.
- **Inside the Computer Society**—Keep up with member activities in the new Inside the Computer Society area.
- **Tech News**—Highlights blogs and new articles, as well as trending papers from the digital library.
- **Education**—View the new IEEE Computer Society Course Catalog featuring a range of educational offerings.
- **Calls for Papers**—Discover opportunities to write and speak with the new Calls for Papers platform.

Discover more at [www.computer.org](http://www.computer.org) today!
Toward a Machine Intelligence Layer for Diverse Industrial IoT Use Cases

Emoji: Lingua Franca or Passing Fancy?

The Online Trolling Ecosystem
Internet of Things

10 Semantic Enablement in IoT Service Layers—Standard Progress and Challenges
KOMAL GILANI, JAHO KIM, JAESEUNG SONG, DALE SEED, AND CHONGGANG WANG

18 Toward a Machine Intelligence Layer for Diverse Industrial IoT Use Cases
JAN HÖLLER, VLASIOS TSIATSIS, AND CATHERINE MULLIGAN

Blockchain

26 Self-Managing Real Estate
NATHAN SHEDROF

27 Blockchain in Developing Countries
NIR KSHTETI AND JEFFREY VOAS

Social Media

31 Emoji: Lingua Franca or Passing Fancy?
GEORGE HURLBURT

38 The Online Trolling Ecosystem
HAL BERGHEL AND DANIEL BERLEANT

Careers

46 CareerVis: Hierarchical Visualization of Career Pathway Data
MINGRAN LI, WENJIE WU, JUNHAN ZHAO, KEYUAN ZHOU, DAVID PERKIS, TIMOTHY N. BOND, KEVIN MUMFORD, DAVID HUMMELS, AND YINGJIE VICTOR CHEN

Departments

4 Magazine Roundup
8 Editor’s Note: Managing IoT Diversity
72 Conference Calendar

Subscribe to ComputingEdge for free at www.computer.org/computingedge.
The IEEE Computer Society’s lineup of 12 peer-reviewed technical magazines covers cutting-edge topics ranging from software design and computer graphics to Internet computing and security, from scientific applications and machine intelligence to visualization and microchip design. Here are highlights from recent issues.

**Computer**

**Hybrid Vehicular Crowdsourcing with Driverless Cars: Challenges and a Solution**

Although vehicular crowdsourcing represents an emerging technology to assist many smart city applications, maintaining sensing data quality is still a challenge. This article from the December 2018 issue of *Computer* considers the challenges and offers a potential solution for a hybrid scenario involving both driverless cars and human-controlled vehicles, within the limited task budget.

**Computing in Science & Engineering**

**Evidence-Based Detection of Advanced Persistent Threats**

This article from the November/December 2018 issue of *Computing in Science & Engineering* presents an approach to the automation of cybersecurity operations centers with
cognitive assistants that capture and automatically apply the expertise employed by cybersecurity analysts when they investigate advanced persistent threats. The goal is to significantly increase the probability of detecting intrusion activity while drastically reducing the workload of the operators.

IEEE Annals of the History of Computing

Oral History of Dov Frohman
Dov Frohman is an Israeli electrical engineer and businessman. In 1970, he invented the Electrically Programmable Read-Only Memory (EPROM), a key enabling technology for rapid development of microprocessor-based systems, from personal computers to industrial controls. Intel founder Gordon Moore called it “as important in the development of the microcomputer industry as the microprocessor itself.” Frohman was also responsible for establishing Intel’s R&D and manufacturing presence in Israel, one of Intel’s most productive and advanced design centers. Frohman is a nationally known figure in Israel. Read more in the October–December 2018 issue of IEEE Annals of the History of Computing.

IEEE Computer Graphics and Applications

Graphoto: Aesthetically Pleasing Charts for Casual Information Visualization
Graphoto is a framework that automatically generates a photo or adjusts an existing one to match a line graph. Since aesthetics is an important element in visualizing personal data, Graphoto provides users with aesthetically pleasing displays for casual line graph information visualization. More specifically, after creating a line graph of the input data, a photo that resembles the input data on the line graph is selected from a photo archive. The authors of this article from the November/December 2018 issue of IEEE Computer Graphics and Applications present a user study to show the effectiveness of Graphoto in terms of data interpretation and aesthetics.

IEEE Intelligent Systems

A Multimodal Approach for the Safeguarding and Transmission of Intangible Cultural Heritage: The Case of i-Treasures
Intangible cultural heritage (ICH) creations include music, dance, singing, theater, human skills, and craftsmanship. These cultural expressions are usually transmitted orally or using gestures and are modified over a period of time, through a process of collective recreation. As the world becomes more interconnected and many cultures come into contact, local communities run the risk of losing important elements of their ICH, while young people find it difficult to maintain the connection with the cultural heritage treasured by their elders. In this article from the November/December 2018 issue of IEEE Intelligent Systems, the authors present a novel holistic approach for the safeguarding and transmission of ICH that goes beyond the mere digitization of ICH content.

IEEE Internet Computing

Considering Jurisdiction When Assessing End-to-End Network Neutrality
Existing solutions designed to assess end-to-end neutrality violations do not consider the normative jurisdictions. The authors of this article from the November/December 2018 issue of IEEE Internet Computing argue that jurisdiction-aware violation detection can be achieved through further steps that can be added to current solutions. As a proof-of-concept, they propose a prototype to expose and discuss the challenges and open issues that need to be faced to consider the normative jurisdiction when assessing end-to-end network neutrality.

IEEE Micro

Image Recognition Accelerator Design Using In-Memory Processing
This article from the January/February 2019 issue of IEEE Micro proposes a hardware accelerator design, called object recognition and classification hardware accelerator on resistive devices, which processes object recognition tasks inside emerging non-volatile memory. The in-memory processing dramatically lowers the overhead of data movement, improving overall system efficiency. The proposed design
accelerates key subtasks of image recognition, including text, face, pedestrian, and vehicle recognition. The evaluation shows significant improvements on performance and energy efficiency as compared to state-of-the-art processors and accelerators.

IEEE MultiMedia

Social Relationship Labeling Based on Multimodal Behaviors and Social Interactions
This article from the October–December 2018 issue of IEEE MultiMedia addresses the social relationship labeling problem by exploiting users’ multi-modal behaviors and abundant social interactions on Twitter. This can effectively alleviate the problem of lacking specific information of following relationships. The experimental results demonstrate the effectiveness of the designed features and different classifiers.

IEEE Pervasive Computing

Pervasive Agriculture: IoT-Enabled Greenhouse for Plant Growth Control
The authors of this article from the October–December 2018 issue of IEEE Pervasive Computing present an Internet of Things (IoT) deployment in a tomato greenhouse in Russia. The IoT-enabling technologies in this deployment are a wireless sensor network, cloud computing, and artificial intelligence. They are to help in monitoring and controlling both plants and greenhouse conditions, as well as predicting the growth rate of tomatoes.

IEEE Security & Privacy

The Good, the Bad, and the Ugly: Two Decades of E-Voting in Brazil
Brazil pioneered the adoption of nationwide electronic voting 20 years ago. However, today its system is outdated in terms of recent properties. The authors of this article from the November/December 2018 issue of IEEE Security & Privacy discuss the system’s organization and transparency mechanisms in the context of security requirements derived from a conventional election.

IEEE Software

Spotify Guilds: How to Succeed with Knowledge Sharing in Large-Scale Agile Organizations
The new generation of software companies has revolutionized the way companies are designed. While bottom-up governance and team autonomy improve motivation, performance, and innovation, managing agile development at scale is a challenge. In this article from the March/April 2019 issue of IEEE Software, the authors describe how Spotify cultivates guilds to help the company share knowledge, align, and make collective decisions.

IT Professional

Autonomous Cars: Social and Economic Implications
One of the major issues with autonomous cars is their future impact on society, as well as on the research community, academia, and industry. As interest in autonomous car technology grows, the social and economic implications of this technology will affect various stakeholders, including its commercialization. In this article from the November/December 2018 issue of IT Professional, the authors critically review and analyze both the economic and social implications of the autonomous car. The significance of these implications will play an important role in the future of autonomous cars among consumers.
IEEE Computer Graphics and Applications bridges the theory and practice of computer graphics. Subscribe to CG&A and

• stay current on the latest tools and applications and gain invaluable practical and research knowledge,
• discover cutting-edge applications and learn more about the latest techniques, and
• benefit from CG&A’s active and connected editorial board.

ADVERTISER INFORMATION

Advertising Personnel
Debbie Sims: Advertising Coordinator
Email: dsims@computer.org
Phone: +1 714 816 2138 | Fax: +1 714 821 4010

Southwest, California:
Mike Hughes
Email: mikehughes@computer.org
Phone: +1 805 529 6790

Advertising Sales Representatives (display)
Central, Northwest, Southeast, Far East:
Eric Kincaid
Email: e.kincaid@computer.org
Phone: +1 214 673 3742
Fax: +1 888 886 8599

Northeast, Midwest, Europe, Middle East:
David Schissler
Email: d.schissler@computer.org
Phone: +1 508 394 4026
Fax: +1 508 394 1707

Advertising Sales Representative (Classifieds & Jobs Board)
Heather Buonadies
Email: h.buonadies@computer.org
Phone: +1 201 887 1703

Advertising Sales Representative (Jobs Board)
Marie Thompson
Email: marie@4caradio.org
Phone: 714-813-5094
The Internet of Things (IoT) pervades many industries, domains, and facets of our lives. Yet, these connected objects are often not standardized and don’t work across different applications. This lack of interoperability is a key challenge in increasing adoption and effectiveness of IoT systems. Two articles in this issue of ComputingEdge present innovations for managing this heterogeneous IoT ecosystem.

The authors of IEEE Internet Computing’s “Semantic Enablement in IoT Service Layers—Standard Progress and Challenges” argue that standardizing a set of common functions across IoT applications would reduce the development cost of IoT devices. They propose a semantic-enabled IoT service-layer platform based on oneM2M standards. IEEE Intelligent Systems’ “Toward a Machine Intelligence Layer for Diverse Industrial IoT Use Cases” offers guidelines to help designers create scalable and replicable IoT systems.

Blockchain technology is being employed in a flood of diverse new applications. In Computer’s “Self-Managing Real Estate,” the author explains how blockchain records could soon be used for buying a house. IT Professional’s “Blockchain in Developing Countries” details ways that blockchain could help fight corruption and promote stability in developing countries.

Next, two articles address phenomena that unite otherwise varied social media platforms. IT Professional’s “Emoji: Lingua Franca or Passing Fancy?” evaluates how people use emojis in digital communication. Computer’s “The Online Trolling Ecosystem” laments the ubiquity of disinformation on social media and calls for a renewed effort to battle its spread.

When it comes to careers, diversity is important. “CareerVis: Hierarchical Visualization of Career Pathway Data,” from IEEE Computer Graphics and Applications, presents a tool for helping young adults explore their many career options.
PURPOSE: The IEEE Computer Society is the world’s largest association of computing professionals and is the leading provider of technical information in the field.

MEMBERSHIP: Members receive the monthly magazine *Computer*, discounts, and opportunities to serve (all activities are led by volunteer members). Membership is open to all IEEE members, affiliate society members, and others interested in the computer field.

COMPUTER SOCIETY WEBSITE: www.computer.org

OMBUDSMAN: Direct unresolved complaints to ombudsman@computer.org.

CHAPTERS: Regular and student chapters worldwide provide the opportunity to interact with colleagues, hear technical experts, and serve the local professional community.

AVAILABLE INFORMATION: To check membership status, report an address change, or obtain more information on any of the following, email Customer Service at help@computer.org or call +1 714 821 8380 (international) or our toll-free number, +1 800 272 6657 (US):
- Membership applications
- Publications catalog
- Draft standards and order forms
- Technical committee list
- Technical committee application
- Chapter start-up procedures
- Student scholarship information
- Volunteer leaders/staff directory
- IEEE senior member grade application (requires 10 years practice and significant performance in five of those 10)

PUBLICATIONS AND ACTIVITIES

*Computer*: The flagship publication of the IEEE Computer Society, *Computer*, publishes peer-reviewed technical content that covers all aspects of computer science, computer engineering, technology, and applications.

Periodicals: The society publishes 12 magazines, 15 transactions, and two letters. Refer to membership application or request information as noted above.

Conference Proceedings & Books: Conference Publishing Services publishes more than 275 titles every year.

Standards Working Groups: More than 150 groups produce IEEE standards used throughout the world.

Technical Committees: TCs provide professional interaction in more than 30 technical areas and directly influence computer engineering conferences and publications.

Conferences/Education: The society holds about 200 conferences each year and sponsors many educational activities, including computing science accreditation.

Certifications: The society offers three software developer credentials. For more information, visit www.computer.org/certification.

2019 BOARD OF GOVERNORS MEETINGS

6 – 7 June: Hyatt Regency Coral Gables, Miami, FL
(TBD) November: Teleconference

EXECUTIVE COMMITTEE

President: Cecilia Metra
President-Elect: Leila De Floriani; Past President: Hironori Kasahara; First VP: Forrest Shull; Second VP: Avi Mendelson;
Secretary: David Lomet; Treasurer: Dimitrios Serpanos;
VP, Member & Geographic Activities: Yervant Zorian;
VP, Professional & Educational Activities: Kunio Uchiyama;
VP, Publications: Fabrizio Lombardi; VP, Standards Activities: Riccardo Mariani; VP, Technical & Conference Activities: William D. Gropp
2018–2019 IEEE Division V Director: John W. Walz
2019 IEEE Division V Director Elect: Thomas M. Conte
2019–2020 IEEE Division VIII Director: Elizabeth L. Burd

BOARD OF GOVERNORS

Term Expiring 2019: Saurabh Bagchi, Leila De Floriani, David S. Ebert, Jill I. Gostin, William Gropp, Sumi Helal, Avi Mendelson
Term Expiring 2020: Andy Chen, John D. Johnson, Sy-Yen Kuo, David Lomet, Dimitrios Serpanos, Forrest Shull, Hayato Yamana
Term Expiring 2021: M. Brian Blake, Fred Douglas, Carlos E. Jimenez-Gomez, Ramalatha Marimuthu, Erik Jan Marinissen, Kunio Uchiyama

EXECUTIVE STAFF

Executive Director: Melissa Russell
Director, Governance & Associate Executive Director: Anne Marie Kelly
Director, Finance & Accounting: Sunny Hwang
Director, Information Technology & Services: Sumit Kacker
Director, Marketing & Sales: Michelle Tubb
Director, Membership Development: Eric Berkowitz

COMPUTER SOCIETY OFFICES

Washington, D.C.: 2001 L St., Ste. 700, Washington, D.C. 20036-4928 • Phone: +1 202 371 0101 • Fax: +1 202 728 9614
Email: hq.ofc@computer.org

Los Alamitos: 10662 Los Vaqueros Cir., Los Alamitos, CA 90720
Phone: +1 714 821 8380 • Email: help@computer.org

Asia/Pacific: Watanabe Building, 1-4-2 Minami-Aoyama, Minato-ku, Tokyo 107-0062, Japan • Phone: +81 3 3408 3118
Fax: +81 3 3408 3553 • Email: tokyo.ofc@computer.org

MEMBERSHIP & PUBLICATION ORDERS

Phone: +1 800 272 6657 • Fax: +1 714 821 4641
Email: help@computer.org

IEEE BOARD OF DIRECTORS

President & CEO: Jose M.D. Moura
President-Elect: Toshio Fukuda
Past President: James A. Jefferies
Secretary: Kathleen Kramer
Treasurer: Joseph V. Lillie
Director & President, IEEE-USA: Thomas M. Coughlin
Director & President, Standards Association: Robert S. Fish
Director & VP, Educational Activities: Witold M. Kinsner
Director & VP, Membership and Geographic Activities: Francis B. Grosz, Jr.
Director & VP, Publication Services & Products: Hulya Kirkici
Director & VP, Technical Activities: K.J. Ray Liu

revised 13 February 2019
Semantic Enablement in IoT Service Layers—Standard Progress and Challenges

We believe that applying semantic technologies to IoT service layer platforms can improve data accessibility, data discoverability, and the ability to extract knowledge about the data. Therefore, this article shows how semantic technologies can be leveraged by IoT service layer platforms.

The Internet of Things (IoT) is significantly growing with an aim to make a connected world by providing numerous opportunities for many industrial sectors and domains such as smart cities, smart factories, and smart homes. Currently, however, IoT applications in these domains are not interoperable with each other. The heterogeneous nature of these applications provide justification for defining a standard way of abstracting vertical data models. IoT data is usually collected from various sources such as sensory devices and/or crowd sensing. The data is often stored in IoT platforms as resources based on different data models. This collection of data can vary in quality and context. Accordingly, a semantic approach—for example, used in the Semantic Web—can provide great agility toward resource representation, sharing information, and inferring new knowledge from data in the IoT on a global scale.

Standardizing a set of common functions (such as registration and discovery) across IoT applications and devices would reduce the development cost of IoT devices. This IoT service layer enables application development independent of the underlying network communication and protocols (such as HyperText Transfer Protocol [HTTP] and Constrained Application Protocol [CoAP]) by abstracting different network technologies. As most IoT service layer platforms simply store IoT data in a non-semantic aware fashion, the meaning of the data cannot be conveyed to IoT applications. Therefore, they are unable to understand the context of the data. Meaningful use of any IoT data requires knowledge about its context such as its geolocation, its

Komal Gilani  
Sejong University

Jaho Kim  
Korea Electronics Technology Institute

JaeSeung Song  
Sejong University

Dale Seed  
InterDigital Communications

Chonggang Wang  
InterDigital Communications
units and its producer. We believe that applying semantic technologies to IoT service layer platforms can improve data accessibility, data discoverability, and the ability to extract knowledge about the data. Therefore, this article shows how semantic technologies can be leveraged by IoT service layer platforms.

Since the way of storing and managing data in IoT platforms is different compared to the Web, this article shows mechanisms of using a data modelling language such as Resource Description Framework (RDF) to semantically describe IoT data, methods of associating IoT data with this semantic metadata, and methods of handling semantic queries to discover meaningful data from IoT platforms. For this purpose, we have designed a semantic-enabled IoT service-layer platform based on oneM2M global IoT standards supporting semantic features such as annotation and discovery.

**SEMANTIC TECHNOLOGIES AND RELATED STANDARDS**

This section explains why semantic technologies are required for IoT service layer platforms and gives an overview of core semantic technologies. Semantic technologies can play a critical role in data and knowledge management for context-awareness in IoT service platforms.

**Ontology**

Ontology represents concepts as objects that have properties and relationships with other objects. An ontology describes linguistic artifacts using a shared vocabulary of basic concepts about a piece of reality. It helps to support semantic exchange and context-driven communications among people and machines by defining shared and common theories.

**RDF and RDF Schema**

RDF is a standard model and language that represents the ontological level of facts about a resource or an individual—for example, types of individuals and their relations, respectively. RDF Schema provides a vocabulary for structuring RDF resources and describing relationships among resources. This includes the modelling of classes (rdfs:Class), the rdf:type property that provides the links of instances to a class, and the rdfs:subclass property, which allows the specification of class hierarchies.

**OWL**

As an ontology language, RDF and RDFS have limited expressiveness, as they have difficulties describing cardinality constraints (e.g., Parking Garage A has more than 10 unoccupied parking spots). Therefore, OWL was introduced to provide greater expressiveness and even support ontological reasoning. OWL offers different sublanguages with different levels of expressiveness and related properties regarding reasoning completeness and time complexity.

**SPARQL**

SPARQL is a query language for interacting with a triple store to process stored RDF triples. SPARQL can support ontological reasoning and semantic discovery. The triple store typically provides an interface to receive SPARQL query requests from a user and to send responses back to the user. Now the question is whether and how these technologies can be leveraged in an IoT service layer platform to support semantic interoperability.
**SEMANTIC-ENABLED IOT SERVICE LAYER**

A common IoT service layer platform is required by the IoT market to facilitate multi-industry IoT applications. The oneM2M Global Initiative is an international partnership project to develop a globally acceptable IoT service-layer standard. The common service layer specified by oneM2M can be embedded into various IoT entities such as end devices, gateways, and servers. It provides various IoT common service functionalities such as device registration, group management, and security and privacy.

The oneM2M service layer provides a means for connecting various IoT devices regardless of their access technologies, collecting data from these devices, and managing the collected data. Through its semantic capabilities, it also supports the annotation of semantic descriptions to oneM2M resources. Figure 1 shows the high-level design of a semantic-enabled IoT service layer platform. In order to support semantic features, IoT service-layer platforms have to support at least three basic features as follows:

- **Semantic annotation**: To achieve data interoperability, the service layer first should be able to support describing the meaning of resources/data. IoT service-layer resources (i.e. data sets) can be annotated with semantic information using standardized ontologies and data structures.

- **Semantic query and discovery**: The platform can support queries from IoT applications based on a semantic query language. When a semantic query is received, the platform executes the query by retrieving semantic information for the targeted resources and processing the discovery query.

- **Semantic mashup**: Like a traditional Web mashup, a semantic mashup is used to compose a virtual IoT resource from more than one IoT resource, which can be other existing virtual resources as well.

In order to provide semantic services to users properly, it is necessary to define common vocabularies, standardized data formats and description rules that can eventually solve the interoperability challenges caused by heterogeneous IoT data. The standard RDF language can be used to describe the semantic information. Also the annotated semantic metadata is then stored by the platform in a new resource designed to accommodate semantic information in an RDF/RDFS format. The metadata can also be stored in a triple store/ontology repository.

![Figure 1. Semantic capabilities in an IoT service layer.](image-url)
SEMANTICS IN ONEM2M STANDARDS

In this section, we describe how the semantic IoT features mentioned in the previous section can be realized in an IoT service-layer platform.

oneM2M Resources for Storing Semantic Information

The two basic logical entities that play a major role in the oneM2M system are an Application Entity (AE) and a Common Service Entity (CSE). In the oneM2M architecture, both CSEs and AEs can reside within different nodes, such as an Infrastructure Node (IN) for a server platform, a Middle Node (MN) for a gateway, an Application Service Node (ASN) and an Application Dedicated Nodes (ADN) for a constrained device. The AE is the logical entity that provides application’s business logic. It is used for hosting sensors, applications, and it resides in the Application dedicated node, which is called AND-AE. On the other hand, the IN-CSE entity is hosted on a server. The CSE functionality is provided for utilization by various AE resources. oneM2M adopted a resource based data model, in which all services are represented as resources. A resource can be uniquely addressed by a Uniform Resource Identifier (URI) and manipulated via create, retrieve, update, delete, and notify operations (CRUD+N).

To enable semantic technologies, the oneM2M service layer defines a <semanticDescriptor> resource, as highlighted in Figure 2. This resource is responsible for storing semantic information related to its parent resource and potentially sub-resources. It is created inside an existing container resource or AE resource of CSE in the oneM2M resource structure. The contents of this resource can be provided based on ontologies. The <semanticDescriptor> resource contains various attributes—that is, ontologyRef for the URI of an ontology, descriptorRepresentation to indicate the format of the semantic information, relatedSemantics to contain the URIs of other related descriptor resources, and descriptor for semantic information itself—to facilitate semantic information management. The <subscription> resource can be added as a child resource by any CSE/AE that expects to receive automatic notifications on the changes of a <semanticDescriptor> resource.

Let’s describe an example of the semantic information management process, where two sensors measure temperature information in different units and a smartphone application makes a discovery request of relevant semantic information. These two sensors are represented as ADN-AE-1 and ADN-AE-2 in an IoT platform server and periodically store measured temperature values in the server. The measured temperature sensor values are stored to a <contentInstance> resource for each sensor reading with a <semanticDescriptor> as its child resource. The <semanticDescriptor> resource is used to store the semantic information about the temperature sensor reading and the measured value. Once the <semanticDescriptor> resource is created, the smartphone application (i.e. ADN-AE-3) sends a semantic discovery request to the IN-CSE,
which contains a semantic filter. Then the IN-CSE will use the semantic filter to discover desired resources. After this the application ADN-AE-3 receives a response in the form of unique resource identifiers. Based on the returned list of unique resource identifiers, ADN-AE-3 can make another request to the IN-CSE to retrieve one or more semantic descriptor resources.

oneM2M Base Ontology

In general, information and operations in each IoT system can be described by ontologies, which provide a vocabulary with a structure. These ontologies (with OWL representations) can be used to support interoperability between different systems via ontology integration or mapping. For this purpose, oneM2M has defined its own ontology called the oneM2M Base Ontology. Various external ontologies from other IoT systems can be mapped to the oneM2M Base Ontology (e.g., by sub-classing and equivalence) so the interworking between the oneM2M system and external systems can be achieved. The oneM2M Base Ontology contains Classes (i.e. sets of individuals) and Properties (i.e. relationships and links between individuals), but no instances since the Base Ontology only supports a semantic description of these entities in the oneM2M architecture.

Semantic Annotation

Semantic annotation, which is the first step toward a semantic IoT system, is a process of adding semantic information to resources in oneM2M IoT platforms so that an annotated resource can be discovered semantically by heterogeneous IoT applications. In the oneM2M system, semantic information is represented using RDF/RDFS (or OWL) as RDF triples. Since the oneM2M system uses a hierarchical tree structure to store and manage its resources, semantic information is added as a special semantic resource. For this purpose, an IoT semantic annotator (IoT-SA) is introduced that runs within the oneM2M IoT system to automatically annotate semantic information for various resources representing sensors/devices registered to the oneM2M system with the following five steps:

1. As inputs to the IoT-SA, users/admins select IoT resource(s) to be annotated from the IoT platform and choose ontology(s) to be used during this annotation.
2. The IoT-SA then parses the given ontology to retrieve its classes and properties. The IoT-SA also retrieves other resources having related semantic information from the platform as candidate resources to establish relationships. The related semantic information is retrieved from <relatedSemantic> attribute, which contains URIs of other linked descriptor resource(s).
3. Users/Admins repeat a process to define semantic information in a triple format (i.e. subject $\rightarrow$ predicate $\rightarrow$ object) based on the given classes and attributes/properties from the given ontology.
4. Selected resources and semantic information are then converted into the defined RDF format and the IoT-SA uploads encoded RDF triples to the <semanticDescriptor> resource under the target resource.
5. The semantically annotated resources can now be discoverable by IoT applications. The updated semantic information can also be seen by users/admins for other purposes.
Semantic Resource Discovery and Semantic Query

One of the key benefits of semantic descriptions is to enable semantic resource discovery. Semantic resource discovery is basically a capability for an IoT application to discover resources based on certain specified characteristics of resources it is interested in. Semantic resource discovery can be achieved by using a SPARQL query. Figure 3 shows semantic discovery procedures in oneM2M. An IoT application is notified of the discovered resources and can retrieve desired resources based on the returned URIs after a semantic query is executed in the oneM2M platform. Specifically, a semantic filter is specified in oneM2M, which is formulated as a SPARQL query and contained in a semantic resource discovery request. An IoT application that wants to discover resources using semantics has to form a semantic query statement using SPARQL based on its needs.

When a SPARQL query is received targeting a specific resource (a.k.a. target resource), the receiver (i.e. IN-CSE in Figure 3) performs Semantic Graph Scoping (SGS) to decide the scope of the SPARQL query execution (i.e. to formulate a RDF basis for executing the SPARQL query). Semantic descriptors which are distributed and are hosted in the IoT platform’s resource structure are collected together to formulate a complete RDF data basis.

Semantic Mashup

Semantic Mashup is a process to discover and collect data from more than one IoT data sources and apply relevant business logic on the collected data to generate meaningful mashup results. For example, let us consider a case where users are interested in a service called “weather comfort index,” which provides and expresses satisfaction level regarding weather conditions. The comfort levels can be calculated based on the temperature and humidity sensors deployed in a specific location together with additional weather conditions; this is actually a mashup process and can be provided as a mashup service by an IoT platform. oneM2M specifies a semantic mashup service, which is implemented via a set of mashup procedures as shown in Figure 4. In order to utilize a mashup service, an IoT application should first discover the corresponding SMJP (i.e. a <semanticMashupJobProfile> resource as defined in oneM2M (Step 1)). A SMJP describes the profile and necessary information required for a specific mashup service such as input parameters, member resources, mashup function, and output parameters. The SMJP resource shall contain <semanticMashupInstance>, <semanticDescriptor> and <subscription> as child resources. Based on the profile described in the SMJP, Originators (e.g. AEs) can create corresponding semantic mashup instances where semantic mashup results will be generated and stored in <semanticMashupResult>. The Mashup Requestor may use <semanticMashupResult> to retrieve the mashup result.
Figure 4. IoT Semantic mashup procedures in oneM2M. Each specific mashup service is described by a Semantic Mashup Job Profile (SMJP) which defines all required elements (e.g. types of input parameters, types of member resources, mashup operations or business logic, etc.) in RDF triples by this mashup service.

Based on the discovered SMJP, the next step is for the IoT application to create a Semantic Mashup Instance (SMI) resource, for example, by giving appropriate input parameters and member resources (Step 2). The SMI resource is used to contain input parameters, member resources, and any generated mashup results. Basically, SMJP provides a guidance on how an SMI shall be created and how the mashup result shall be calculated. The third step is for the IoT platform (i.e. CSE in oneM2M) to discover and collect original data from each member resource (e.g. via semantic resource discovery procedures) (Step 3). After the data is collected from the identified member resources (i.e. data sources), the IoT platform calculates the mashup result according to the business logic as described in the SMJP (Step 4). The generated semantic mashup result is stored in the SMI, which can be retrieved by the IoT application or other entities (Step 5).

CONCLUSION

There is a strong need to resolve the interoperability issue in the IoT service layer using semantic technologies inspired by semantic Web. This article described a semantic-enabled IoT service layer architecture based on the oneM2M global IoT service layer standards. In this architecture, semantic descriptor resources are introduced to represent semantic information in RDF triples. This semantic descriptor resource allows an IoT service layer or an IoT application to annotate existing IoT resources/data with additional semantic information using selected ontologies and RDF/RDFS. The added semantic information is then leveraged for semantic filtering/discovery and semantic mash-up.

The proposed semantic-enabled IoT architecture also supports a semantic repository to maintain all semantic information in a centralized triple store. Then, a SPARQL query can be executed directly on the triple store against the semantic information stored there. Future work includes advanced semantic annotation with other data models, information synchronization between oneM2M service layer resource structure and the triple store, distributed semantic analytics and other functions, as well as interoperability with other standards.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No. B0184-15-1003, No. B0184-15-1001)
REFERENCES


ABOUT THE AUTHORS

Komal Gilani is an MSc student at Sejong University. Her research interests include IoT semantics, smart cities, and industrial IoT. She has a BS in computer science from Arid Agriculture University Rawalpindi. Contact her at G2017SFE052@sju.ac.kr.

Jaeho Kim is a principal engineer at Korea Electronics Technology Institute. His research interests include IoT, smart cities, and digital twin technologies. He has a PhD in electrical and electronic engineering from Yonsei University. Contact him at jhkim@keti.re.kr.

JaeSeung Song is an associate professor at Sejong University. His research interests include IoT semantics, software testing, and industrial IoT. He has a PhD in computing from Imperial College London. Contact him at jssong@sejong.ac.kr.

Dale Seed is a principal engineer at InterDigital Communications. His research interests include IoT protocols and services. He has a MS in computer science and engineering from the Pennsylvania State University. Contact him at dalseed@interdigital.com.

Chonggang Wang is a principal engineer at InterDigital Communications. His research interests include IoT, blockchain and distributed ledger technologies, and AI-powered future networking and computing. He has a PhD in computer science from Beijing University of Posts and Telecommunications. Contact him at chonggang.wang@interdigital.com.
The Internet of Things (IoT) has moved beyond the hype, and today we see promising applications materializing and industries transforming through well-known digitalization as well as servitization—that is, delivering a service as an integral part of a product. This is evident in the increasing number of physical industry assets represented and manipulated in both the digital and physical worlds and in the fact that the business models for physical and digital assets are converging toward service as opposed to product sales.

IoT can be used in many industry sectors with numerous benefits. Cost optimization and environmental efficiency are just two factors driving this expansion. Examples of IoT applications include predictive maintenance and condition-based monitoring, which are mainly used in industrial settings. However, the envisioned IoT applications are so diverse and include such a broad spectrum of technologies that system designers need design support tools and guidelines. This article provides a few of these design guidelines in the form of architecture and design patterns to enable scalable and replicable solutions rather than point solutions stemming from point problems. As a result, this article focuses on sharing our experience with the analysis of several IoT use cases and a machine intelligence framework that combines knowledge of solution design for them.

**Approach**

The main goal of building replicable solutions can be likened to the goal of formulating a reference architecture that encompasses and encodes the knowledge of numerous solution architectures. In turn, according to Nick Rozanski and Eóin Woods, a solution architecture can be described as a set of architecture views or blueprints, each addressing the concerns of a specific stakeholder.¹

In generating a reference architecture one typically follows the design process for a single stakeholder concern that is typically expressed with one or more use cases. The process is then repeated for all possible stakeholder concerns. In the end, all solution architectures are combined in a union.

In this article, we follow the reference architecture process for a subset of stakeholders (end users); therefore, the union of different solution architectures is a partial version of a reference architecture called the *machine intelligence layer* for diverse industrial IoT use cases.

---

Figure 1. Structuring the problem and solutions domains (MI: machine intelligence, SLO: service-level objective).

Framework. Figure 1 illustrates this process, referred to as **structuring the solution domain**. Other types of stakeholders are developers and service providers. The resulting machine intelligence framework with a different stakeholder will be somewhat different from that of an end user. However, following the same methodology outlined here, frameworks for other stakeholders can be generated.

Because of the huge number of available use cases, with new ones arising all the time, following the design process to generate a single solution architecture is not tractable. Therefore, we also structure the problem domain or the potential use cases. For the purposes of this exercise, we limited the studied use case sources to oneM2M,² the Industrial Internet Consortium (IIC) use cases and testbeds (www.iiconsortium.org), and National Institute of Standards and Technology big data.³ In the context of this article, structuring the problem domain (Figure 1) means that individual uses cases are grouped into use case families or patterns. We have used the use cases’ structure to limit the number of times we applied the design process for developing solution blueprints. As a result, instead of generating solution blueprints from each specific use case, we generated architecture blueprints for a representative use case from a family of use cases or a use case pattern. We then iterated the design process over all the identified use case patterns.

**Structuring the Problem Domain: Use Case Patterns**

We have studied about 100 use cases from different sources, which have varying degrees of detail and are classified or grouped according to market-related terms. The most typical of these are the (market) sector and (market) vertical that a use case belongs to. Although not well-defined in the literature, we have followed the IIC definition and taxonomy.⁴ According to the IIC, a sector (such as healthcare) is a logical group of related verticals (for example, hospitals), and a vertical is a market in which vendors offer goods and services that meet a particular set of usage, technical, or regulatory requirements.

We focus our study on the end user as the main system stakeholder as stated earlier. The end user formulates concerns that are concretely described in use cases. These are grouped logically into (use case) applications apart from their market classification into sector and vertical. Our assumption is that a sector contains multiple verticals, each containing multiple applications and multiple use case instances of different use case types. We refer to a use case instance as simply a “use case” (see Figure 2). In turn, each use case expresses a stakeholder concern requiring a desired (technical) system functionality or set of characteristics.

Although this classification is done mainly from a market perspective, we aimed to resuffle the same set of use cases into another set of groups that emphasize the technical characteristics. We call these groups (technical) **use case patterns**. We also identified a set of technical characteristics based on which the different (technical) use case patterns can be described; however, for brevity we omit the technical characteristics from the pattern definitions.

Through our research, we have identified seven use case patterns.
Massive monitoring. This use case pattern involves numerous sensors deployed across a large geographical area. Data is collected over a period of time for bulk batch or stream analysis. Data analysis aims to find trends, detect anomalies or abnormal situations, or simply learn the behavior of the monitored asset or phenomena. Typical examples include environmental and climate monitoring and pollution monitoring.

Asset management. This pattern involves managing physical assets that are well-defined and confined. Example assets include a building, vehicle, piece of industrial machinery, turbine, or human patient. Typical management aspects are to optimize the asset’s operation, perform diagnostics, or do predictive maintenance.

Logistics. A typical logistics scenario can be described as the process of coordination, management, and orchestration of a collection of tasks in a workflow to achieve a set of goals (such as time or cost optimization) by making efficient use of the available resources.

Typical examples include fleet management, supply-chain optimization, and pickup-delivery services.

Remote operations. Remote operation generally refers to the control and operation of a system or equipment from a remote location either by humans or software. In this case, the remotely operated system or equipment cannot or is not designed to operate completely autonomously to accomplish the task. Typical examples include remote mining, remote-controlled vehicles and drones, and remote surgery.

Robots and autonomous machines. This use case pattern covers the operations and management of partially or fully autonomous systems such as robots, vehicles, and drones. Such systems are often described as cyberphysical systems (CPS), which integrate the dynamics of the physical processes with those of the software and networking. Typical examples include static or mobile factory floor robots and autonomous vehicles.

Infrastructure monitor and control. This use case pattern covers management of large-scale industrial or extended infrastructures that need to be monitored and controlled. Examples include a transportation infrastructure such as a national road network, a utility infrastructure such as the electric grid, oil and gas pipelines, and street lighting.

Device swarms. This use case pattern covers devices and systems that operate autonomously with a simple set of rules and no central intelligence, and form peer-to-peer groups to collectively reach a common goal. Typical examples include microgrid producers and consumers and zero-trust computing applications such as home automation.

Structuring the Solution Domain

The solution domain consists mainly of architectural blueprints that include a few main types of system components. This domain encompasses devices, connectivity, cloud and distributed computing, machine intelligence, and mechanisms for visualizing information and integrating applications to an enterprise environment. These blueprints typically concretely express the functional components of a resulting solution. However, a system also typically consists of a set of non-functional characteristics. Through our research we have identified a core set of functional and nonfunctional characteristics that can be grouped into different perspectives: data and information perspectives (multimodality of IoT data, the need for insight and analysis-driven functions, knowledge representation, cognition, and so on); control perspectives, meaning whether a control functionality exists (sensing and actuation in feedback loops, workflow/
process-driven); and general characteristics (locality, timing criticality, safety, and security).

**A Framework for Distributed Machine Intelligence for Industrial Use Cases**

Our approach is to aggregate solution blueprints for the identified recurring use case patterns, thus arriving at our desired framework. As mentioned, our focus in this article is distributed machine intelligence functionality supporting the diversity of IoT use cases we have identified.

For the sake of brevity, we have left out two important, but for the objective of this article, secondary considerations. The first is the general need for distributed processing of machine intelligence logic, and the second is lifecycle management of the solution.

IoT data processing and decision making is generally a highly distributed capability. The need for distributed processing in IoT comes from different requirements. Data volumes, cost, performance and latency, autonomous local asset operation, robustness, and safety around IoT asset operation are the main requirements. IoT distributed processing extends beyond the datacenter to constrained IoT devices, and the resulting heterogeneity needs to be managed to meet service-level agreements for the applications.

Lifecycle management includes operational aspects such as the type of logic to deploy and location of the deployment, ensuring necessary robustness, and trust and security. In addition, the system should be adaptive and cognitive to handle changing external requirements or changing contexts. For instance, the system should ideally detect the need to change controller parameters that might be wrongly set or continuously do model training.

**Machine Intelligence Functional Domains**

The machine intelligence framework provides the functionality needed to realize the use case applications relevant to end users. As such, it processes IoT data from various sources, derives and executes control operations to manipulate the asset or infrastructure via actuators, and maintains a cognitive knowledge base related to the assets. An objective of the framework is to partition functionality into application-independent building blocks. These can then be interconnected to realize the use case applications according to a service-oriented paradigm, lending themselves to microservices implementations.

The main functional domains, as shown in Figure 3, include the capability to manage data and relevant IoT resources (sensing, actuation, and identification) and process data and extract information (analytics or machine learning); various types of control and execution (controllers or planning tools); and capabilities to manage and represent knowledge relating to the assets and the system. Figure 3 shows the functional domains as well as some of the main interfaces between them.

**Data and Resource Processing**

By data and resources we mean sensor data, actuator services, and their representations. Individual sensor data and actuator control is the raw fabric for interacting with the physical world. Sensor data includes individual data items and events and datastreams from a single sensor. Resources are abstractions of sensors and actuators in the system.

Massive-scale IoT deployments require some key considerations in the collection phase. Data can be received as an event stream with varying speed, volume, and dynamicity over time. Information creation, attribute validation, and verification are necessary steps in the data collection. Data can be received asynchronously or synchronously, depending on the type of application at hand.

Data management, curation, and resource management are crucial in IoT systems and comprise the following steps. First, data is collected and distributed. Then, data and resources are modeled to capture heterogeneity (structured, unstructured), high distribution, large size (number of data sources, streams, and actuator end points), and semantic annotation describing meaning of the endpoint capabilities. Resources need the proper annotation, describing such attributes as meaning, origin, and quality. Also, transmitted data needs to be filtered to the application needs. IoT data, when needed, will require distributed storage, taking into account factors such as cost and storage capacity.

Real resources require appropriate abstract representations and management in an IoT system, for example, as a representation of a datastream or as an aggregation of sensor data. Resource abstraction implies a level
of indirection requiring a resolution function that dynamically maps between the resource representation and the real resource.

**Insight Generation**
Forecasting is a key issue in the prominent IoT use case of predictive maintenance, which is used to determine the health of a piece of machinery and understand when any maintenance might be needed.

Forecasting involves predicting new outcomes based on previously known results. Depending on the IoT use case, different forecasting timeframes apply. For example, trajectory forecasting of moving objects can be real time, whereas machine degradation is more long term. Forecasting can be data driven or model driven depending on the problem requirements. Model training is a necessity and can be based on training sets or via reinforcement learning. Typical forecasting models can be statistical or neural networks-based, Bayesian or non-Bayesian, linear or nonlinear, parametric or nonparametric, univariate or multivariate.

Sensor fusion is another technique. In general, fusion concerns combining data and information from diverse sources so that the resulting information is more accurate than if one had relied on a single source. An example is the localization of an object that can rely on a combination of ultra-wide band (UWB) transponders, camera detection, and contextual information sensed by the object itself, and when fused provide a much higher degree of location accuracy.

**Knowledge Management**
Knowledge management involves representing, modeling, structuring, and sharing knowledge about a physical asset or infrastructure. Knowledge is the collected set of data, inferred knowledge and insights, and control capabilities of the asset. The knowledge can further be structured so the different data, insights, and control capabilities can be directly mapped to the asset’s real-world structure, thus becoming a proper digital representation of the asset.

Knowledge is generally of two types: declarative knowledge (also referred to as propositional knowledge) and procedural knowledge (also referred to as imperative knowledge). Declarative knowledge describes what an entity is and how it is structured and formally expressed using ontologies. Procedural knowledge describes how an entity behaves, for example, in response to stimuli; and the formal description format is typically via state machines.
Knowledge is captured and made available in what can be referred to as a knowledge base, which is manifested by a set of ontologies. Typical ontologies in IoT are not only for the actual real-world model of the asset and expert knowledge but also knowledge about system and application objectives, such as key performance indicators (KPIs), a work order, task plans, and constraints of the IoT system itself. The real-world model is typically a hierarchical or graph structure.

Across domains in IoT, semantic interoperability is essential for achieving many business applications. Semantic interoperability enables data and information to be shared across domains and understood by systems without needing manual interpretations on top of technical details or protocol and syntactic interoperability. Semantic interoperability requires mapping methods that can be predefined or self-learning. The latter requires algorithms that consider structural, terminological, and semantic differences and similarities.

Object Management

Object management involves identifying, localizing, and cataloging physical assets that are handled by the IoT system. This is important for some types of use cases, such as logistics involving transported goods or localization of tools on a factory floor.

Object identification is possible using various techniques, such as tags based on optical or radio technologies (for example, QR codes or RFID tags). The purpose is to uniquely identify and name objects. A resolution infrastructure is usually in place to find information about the object. A prominent example is electronic product code information services (EPCIS).

Object localization needs to be tailored to the IoT needs and deployment scenarios (such as indoor or outdoor environments). Typical indoor localization technologies include video or image processing, Bluetooth beacons, use of Wi-Fi access points, or UWB ranging. Outdoors, GPS-based localization is typically relied on. For any localization solution, the required accuracy, size of area covered, and real timeliness of location must be considered.

A catalog function can also be required. This function works as a repository of all assets of interest and includes other properties of the asset. EPCIS is an example.

Controllers

Control is a core automation point in any IoT system involving actuators. Control software commands the assets’ desired behavior. Common to all controllers is the deterministic behavior of controlling operations based on input from an a priori desired and defined operational behavior. The use of different controller types is based on functional and nonfunctional characteristics meeting application needs.

Whereas many control systems in robotics and other real-world continuous and industrial systems use proportional, integral, and derivative (PID) controls, other IoT use cases, such as home automation, often use rule-based systems for event-driven control.

A PID controller is a control loop feedback mechanism using a mathematical function that takes the deviation between the desired state and the measure state as input for control. Proportional control means that proportional feedback of the deviation is provided to determine the control value. A derivative part of the deviation dampens the error. An integral part of the deviation provides errors to be removed over time. Examples include inverse kinematics for robot control and temperature control of a fluid system. This requires knowledge about the physical behavior and properties of the asset controlled.

Rule-based controllers are based on a set of predefined rules that are trigger-action pairs, where a trigger is a condition and an action is a predefined workflow typically containing commands to the devices or related services—for example, following the simple logic of “if this, then that.”

In sufficiently complex, dynamic, and nondeterministic situations one can enhance the usability and maintainability of both PID and rule-based control systems by making them use task planning technologies to help infer the actions to be taken.

Task Planning

Task planning can be defined as the process of generating a sequence of actions with certain objectives. Planning
can be applied to a variety of problems such as route planning of autonomous vehicles, optimization of logistics flows, and automation of field personnel.

The planning problem is normally represented by three key elements—states, actions, and goals. State identifies the model of the world, actions represent different operations that affect the system's state, and goals are states to achieve or maintain. Deriving the task plan is to take the current state, the desired state and the possible actions and from that generate a plan as a sequence of possible or proposed actions. A plan can also be a partially ordered list of tasks. One possible way to perform task planning is using AI planners, where the world and the problem are modeled using a planning domain definition language (PDDL).

Multiobjective Optimization

Automating complex system operations by leveraging data-driven strategies designed to analyze alternatives under multiple conflicting views or KPIs is challenging. First, KPI evaluations are not always reliable and might be subject to changes over time; second, the costs incurred in adapting solutions under operation must be accounted for. In such cases, it is difficult to track how the underlying tradeoffs (such as return versus risk or throughput versus cost) will evolve over time, and decision-making preferences are hard to elicit and represent computationally. In the absence of clear preferences and priorities over the KPIs, general problem-solving strategies and architectures must be designed for automating general data-driven multiobjective optimization (MOO) systems under uncertainty.

MOO can play a key role in applications where conflict resolution is expected. For instance, in supply-chain control applications, the proposed system can monitor the profitability for the whole chain as well as the overall product shortage risk. Those two KPIs are clearly in conflict as optimization at an extreme for one results in a risk for the other.

A key difference between task planning and optimization is that in the latter does not assume that desired goal states will be input by the system stakeholders. This stems from the fact that it can be impossible for humans to cope with the underlying complexity of explicitly specifying goal states while simultaneously fulfilling all service-level objectives (SLOs). In such cases, it is possible to leverage simulation-based MOO to automatically explore the space of all candidate goal states that not only fulfill all SLOs but actually surpass them and deliver outstanding performance.

Service Level Objectives and Workflow Management

The end user's interests in the system can be specified as a set of high-level, quantifiable performance metrics by SLOs and workflow orders. SLOs are translated into KPIs, which are deemed critical for verifying service execution and detecting deviations from SLOs. KPIs can further be broken down to needed insights and, together with workflow orders, the intentions or actions of the system. The insights and actions can then be used to define the needed sensor data and actuator controls. For instance, in a logistics use case, a workflow order can request that a number of products be delivered to a certain subset of retailers within a specified deadline to keep shortage risk under the agreed levels.

The KPIs and workflow orders encapsulate information that allows the extraction of inputs to task planning, controllers, and MOO, which also includes the necessary information from the insight generation functional domain. For task planning, the extracted inputs should correspond to goal states that can be used to compute an appropriate plan.

For controllers, workflow orders might specify new set levels of parameters or rules. For multiobjective optimizers, workflow orders should specify a set of KPIs to be balanced by automated tradeoff analysis to comply to overall service objectives as well as mitigating conflicts.

IoT is about the digital representation of the physical world to enable the digitalization and servitization of physical assets or entities of interest. Since the application spread in today's IoT is wide and is typically structured in market-oriented groups, a system designer needs IoT system design patterns to assist in designing for scalable and replicable solutions. The work presented here provides a generic blueprint for designers to jumpstart the design process of an unknown use case.

References

2. ETSI, oneM2M Use Case Collection, ETSI TR118501v1.0.0, tech. report, 2015; www.etsi.org/deliver/etsi_tr /118500_118599/118501/01.00.00_60 /tr_118501v01000p.pdf.
expected. For instance, in supply-chain architectures must be designed for au-
general problem-solving strategies and hard to elicit and represent computa-
tions are not always reliable and might be applied to a variety of prob-
cases, it is possible to leverage simu-
KPIs is challenging. First, KPI evalua-
tions by leveraging data-driven strat -
the planning problem is normally
controllers, and MOO, which also
extraction of inputs to task planning,
shortage risk under the agreed levels.
the workflow orders
request that a number of products be
For instance, in a lo-
tentions or actions of the system. The
together with workflow orders, the in-
SLOs are translated into KPIs, which
The end user’s interests in the sys-
states that not only fulfill all SLOs but
explore the space of all candidate goal
ation-based MOO to automatically
the insight generation func-
tions that can be used to
extracted inputs should correspond
tional domain. For task planning, the
from the insight generation func-
4. R.A. Martin and A. Soellinger, “The
2. ETSI, 
Software
1. N. Rozanski and E. Woods,
References
The work presented here provides a
generic blueprint for designers to
for scalable and replicable solutions.
design patterns to assist in designing
8. EPC Information Services (EPCIS)
Standard, release 1.2, GS1, 2016; http://
www.gs1.org/sites/default/files/docs/epc
EPCIS-Standard-1.2-r-2016-09-29.pdf.
Jan Höller is a research fellow at Ericsson Research, Sweden. His research interests include industrial Internet of Things systems, digital transformation, and machine intelligence in autonomous systems. Höller has an MSc in engineering physics from Lund Institute of Technology. He serves on the Board of Directors of the IP for Smart Objects Alliance. Contact him at jan.holler@ericsson.com.

Vlasios Tsiatsis is a senior researcher at Ericsson Research, Sweden, and an Internet of Things architect. His research interests include IoT, cloud, and analytics security. Tsiatsis has a PhD in electrical engineering from the University of California, Los Angeles. He is a member of IEEE and ACM. Contact him at vlasios.tsiatsis@ericsson.com.

Cathy Mulligan is a research fellow at Imperial College London and vice chair for ETSI Industry Specification Group for Context Information Management. Her research interests include digital technologies and its impact on industrial structures. Mulligan has a PhD in engineering from the University of Cambridge. She is a member of IEEE, IET, and ACM. Contact her at c.mulligan@imperial.ac.uk.

This article originally appeared in IEEE Intelligent Systems, vol. 32, no. 4, 2017.
Self-Managing Real Estate

Nathan Shedroff, Seed Vault Ltd.

Blockchain is poised to make a sea change in just about every industry and business, even real estate.

Remember that folder with all the important papers related to your house? You know, the one with your mortgage and insurance documents, the foundation repair bill, the estimate to redo the electrical for that home theater, and the map showing that your neighbor’s driveway is actually 2 ft. on your side of the property line? Very soon, a new technology called blockchain might allow the house itself to track what happens to it, so that you (and subsequent owners) don’t miss anything. In fact, blockchain is poised to make a sea change in just about every industry and business.

Blockchain is an open, peer-to-peer (meaning shared) ledger of transactions. It’s accounting, but the ledger doesn’t live in a central place—it’s distributed and supported by everyone’s systems. So to authenticate a transaction, more than one node has to “see” the transaction occur and agree that it’s accurate. Only then is the transaction added to the ledger. Once the block (transaction page) is filled, it’s permanently recorded, and a new block is started. The block can’t be changed once it’s verified—it’s part of the permanent record. This means that every single transaction can be retrieved forever, creating a radical new kind of transparency. And everyone should be considering how such transparency will affect their business and their industry.

But back to your house. Blockchain technologies are governed by software code called smart contracts. Think of this code as self-running rules that automate all of the processes. If designed correctly, a business can practically manage itself based on the self-running code.

Therefore, blockchain technology could enable your house to manage its own transactions. It’s not going to call the plumber—yet—but it could manage all of the those documents. Sometime in the next 10 years, rather than relying on a real estate agent or current owner to share the provenance of a property you might be able to submit a query and get a report on every transaction for a house or condo: every repair, change of hands, valuation estimate, tax assessment, redistricting, construction document, and lien. You’ll no longer wonder whether that house is built on an old burial ground or whether the agent failed to mention the meth lab the past owners operated. The blockchain record will tell you, and you’ll be able to trust it. And the house will continue collecting and recording this information through a network of blockchain-enabled services, including identity, storage, and transaction tokens. Your home and its relationships will be recorded for all to see, which might be liberating or creepy—or both.

We don’t normally think of objects as having agency (being entitled to act and make decisions for themselves), though we all know they have a history. This first step is an extension of that history: making it visible to all. The next step—giving objects autonomy—is the subject for another column entirely.

Nathan Shedroff is executive director of the non-profit Seed Vault Ltd., an independent bot blockchain community. Contact him at nathan@nathan.com.

This article originally appeared in Computer, vol. 51, no. 1, 2018.
Blockchain in Developing Countries

A large portion of the population in the developing world can benefit from blockchain technologies. According to the ICT Facts and Figures 2017 report, 42.9 percent of households in developing countries have Internet access. This percentage is rising quickly due to the increasing affordability and usability of smartphones. It can be argued that in many ways, blockchain has a much higher value proposition for the developing world than for the developed world. Why? Because blockchain has the potential to make up for a lack of effective formal institutions— rules, laws, regulations, and their enforcement. In this article, we will discuss key concerns regarding institutions in the developing world and evaluate the potential use of blockchain to address them.

PROPERTY RIGHTS

According to a 2011 UN report, weak governance led to corruption in land occupancy and administration in more than 61 countries. Corruption varied from small-scale bribes to the abuse of government power at the national, state, and local levels.

Enforcement of property rights incentivizes investment and provides resources to avoid poverty. Agreed-upon property rights allow entrepreneurs to use the assets as collateral and thus increase their access to capital. However, a large proportion of the poor lack property rights.

Around 90 percent of land is undocumented or unregistered in rural Africa. Likewise, a lack of land ownership remains among the barriers to entrepreneurship and economic development in India. One estimate suggests that more than 20 million rural families in India do not own land and millions more lack legal ownership of the land where they have built houses and worked. Landlessness is arguably a more powerful predictor of poverty in India than caste or illiteracy.

In addition, according to the United States Agency for International Development (USAID), only 14 percent of Hondurans legally own their properties. Among those properties that are occupied legally, only 30 percent are registered.

It is not uncommon for government officials to alter titles of registered properties, and there are cases where government officials have allocated properties with altered titles to themselves. Bureaucrats have reportedly altered titles and registered beachfront properties for themselves, and have allegedly accepted bribes in exchange for property titles. Citizens often lack access to records, and those records that are accessible might provide conflicting information. Property owners are often unable to defend themselves against infringement of property usage and mineral rights.
Blockchain can reduce friction and conflict, as well as the costs associated with property registration. It is possible to do all or most of the processing using smartphones. Given this, it is encouraging that various initiatives have been undertaken. The US-based platform for real-estate registration, Bitland, announced the introduction of a blockchain-based land registry system in Ghana, where 78 percent of land is unregistered. There is a long backlog of land-dispute cases in Ghanaian courts. Bitland records transactions securely, with GPS coordinates, written descriptions, and satellite photos. This and similar processes are expected to guarantee property rights and reduce corrupt practices. As of mid-2016, 24 communities in Ghana had expressed interest in the project. Bitland is planning to expand to Nigeria in collaboration with the OPEC Fund for International Development.

The bitcoin company BitFury and the Georgian government signed a deal to develop a system for registering land titles using blockchain. Currently, to buy or sell land in Georgia, the buyer and the seller must use public registry. They will pay between $50 and $200, depending on the speed with which they want the transaction notarized. This pilot blockchain project will move the registry process to blockchain. The costs for the buyer and the seller is now expected to be in $0.05-$0.10 range.

In 2017, India’s Telangana and Andhra Pradesh states announced plans to use blockchain for land registry. Telangana started a land registry pilot project in the capital city of Hyderabad. It was reported in September 2017 that a complete rollout of the program in Hyderabad and nearby areas would take place within a year. In October 2017, the Andhra Pradesh government collaborated with a Swedish start-up, ChromaWay, to create a blockchain-based land registry system for the planned city of Amaravati.

CONTROLLING CORRUPTION

Blockchain creates a tamper-proof digital ledger of transactions and shares the ledger, thus offering transparency. Cryptography allows for access to add to the ledger securely. It is extremely difficult—if not impossible—to change or remove data recorded on a ledger. With this feature, blockchain makes it possible to reduce or eliminate integrity violations such as fraud and corruption while also reducing transaction costs.

As an example, the use of fake export invoices to disguise cross-border capital flows has been pervasive in China. During April to September of 2014, $10 billion worth of fake trade transactions were discovered. Major fraud cases occurred at the Qingdao port, where companies had used fake receipts to secure multiple loans against a single cargo of metal. The Qingdao incident involved 300,000 tons of alumina, 20,000 tons of copper, and 80,000 tons of aluminum ingots. As a result, Chinese banks charge higher interest rates and are less likely to offer collateral-based financing. Blockchain can thwart such scandals.

Blockchain also makes it possible to generate smart (“tagged”) property and control it with smart contracts. Examples of such properties include physical property (car, house, container of metal) as well as nonphysical property (shares in a company). Blockchain-based smart properties only undergo actions based on the information published in a smart contract. If property is being used as collateral, the smart contract might not allow the owner to extend the same property as collateral or security to another bank. Thus, the process of verifying collateral prior to the loan being made is greatly simplified for custodians. Here, a trusted trading system is created for smart properties, making credit more readily available and cheaper.

DISADVANTAGED GROUPS

Blockchain might also help refugees and displaced persons. Current systems that offer aid to refugees and displaced persons suffer from inefficiency, fraud, and gross misallocations of resources. For instance, fees and costs account for up to 3.5 percent of an aid transaction. Moreover, an estimated 30 percent of development funds fail to reach the intended recipients due to third-party theft, mismanagement, and other problems.
Various blockchain-based solutions to such problems now exist. For example, blockchain can empower donors by ensuring that their donations reach the intended recipients. For instance, donors can buy electricity for South African schools using bitcoin. A blockchain-enabled smart meter makes it possible to send money directly to the meter, and there are no organizations involved to redistribute funds. Donors can also track the electricity being consumed by schools and calculate the amount of power their donations provide. This program was launched by South African bitcoin startup Bankymoon via a crowdfunding platform.

The UN’s World Food Program (WFP) has used blockchain to help refugees. Money is paid directly to the merchants instead of the recipients. No banks are involved—beneficiaries receive goods directly from the merchants. In early 2017, WFP launched the first stage of what it calls Building Blocks, giving food and cash assistance to needy families in Pakistan’s Sindh province. An Internet-connected smartphone authenticates and records payments from the UN agency to food vendors, ensuring the recipients got the expected help, the merchants got paid, and the agency could keep a watchful eye on the money.

Starting in May 2017, WFP started distributing food vouchers in Jordan’s refugee camps by delivering cryptographically unique coupons to participating camp supermarkets. Supermarket cashiers were equipped with iris scanners to identify the beneficiaries and settle payments (UN databases verify biometric data about refugees). Building Blocks’ ledger records the transactions on a private version of ethereum (a cryptocurrency). WFP reported that by October 2017, it had distributed $1.4 million in food vouchers to 10,500 Syrian refugees in Jordan. WFP expects blockchain to reduce its overhead costs from 3.5 percent to less than 1 percent and to hasten aid to remote or disaster-struck areas (where ATMs might not exist or banks are not functioning normally). Blockchain currency can even replace scarce local cash, allowing aid organizations, residents, and merchants to exchange money quickly and electronically.

SUMMARY

Blockchain will positively affect developing countries: it can help reduce fraud and corruption and increase legal property titles, which provides entrepreneurial initiatives to the world’s poorest. It can also help financial transactions take place more quickly and ensure that aid is distributed with a smaller chance of theft and fraud.

REFERENCES


**ABOUT THE AUTHORS**

**Nir Kshetri** is a professor of management at the Bryan School of Business and Economics, University of North Carolina at Greensboro. Contact him at nbkshetr@uncg.edu.

**Jeffrey Voas** is an IEEE Fellow. Contact him at j.voas@ieee.org.
Emoji: Lingua Franca or Passing Fancy?

Many express awe over the creative use of emoji. Others disdain 😞 the perceived dissolution of proper English.

Yet others fully embrace 😊 the free expression of emotions that emoji enable 😊, while some are infuriated 😡 by the wanton devolution of culture exemplified by such primitive drawings. Many, however, remain indifferent.

Emoji are not new. The humble emoji, as a pictogram, a pictorial representation of an object, or an ideogram (a symbolic representation of a more abstract concept), enjoys a rather long heritage. One could argue that symbolic visualization extends back to prehistoric cave drawings.1

Legendary tribal norms, however, were mostly conveyed by aural means. Starting around 3200 BC, specially selected and educated scribes began etching Egyptian hieroglyphics into stone depicting nobility, conquests, and mysticism. Around the same time, the Sumerian cuneiform emerged as a pictographic script. It morphed over centuries to a more symbolic form of expression. Chinese calligraphy originated around 1200 BC as pictographic script. Around the same time, early pictograms predated the Aztec culture in Mesoamerica with its distinctive illustrative style of writing. In medieval times, educated monks scribed illuminated manuscripts, combining symbolic visual artistry with the written word to preserve religious history on paper. The hybrid visual rebus, also mixing emoji-like illustration with words, often in the form of visual puzzles, also enjoyed growing popularity.

Along the way, symbolic alphabets eventually enabled printing. Once in print, linear strings of symbols rapidly led to universal literacy-based education. After Gutenberg in 1450, knowledge became reproducible, portable, and essential. Printing, the very notion of linearity as reinforced by Newtonian physics, eventually led to production lines. Industrialized economies followed. Eventually, radio reopened aural space and television re-opened visual kinetics. In a short period of relative time, attention shifted from mass production to mass media starting around 1900 and culminating in the dynamic World Wide Web by 1999.

In response, post-modernism elevated consumerism to artfulness in the later 1900s. The smiley face pin became an overnight cultural icon in 1963. Around 1982, emoticons, the use of fonts to form facsimiles of human expression, became vogue. These font combinations conveyed emotion into otherwise dull texts. Influenced by Japanese graphics, Shigetaka Kurita first created the emoji in 1999. It burst quickly onto the Internet. Figure 1 loosely traces this long tail of visual language in human communications, leading to today’s comic-inspired emoji.
While emoji have a long heritage, they are also clearly a product of the digital age. They are now largely standardized into some 1,644 icons in the Unicode Emoji Version 11.0 released on June 5, 2018 (https://unicode.org/emoji/charts/full-emoji-list.html). Thus, they can be quickly produced via keystroke with no need for hand drawing. Using pictograms and ideograms, they frequently convey both thought and emotion. Emoji even follow loose syntax and grammatical rules. This suggests some degree of competence, perhaps even emoji literacy, to become a truly effective emoji communicator. This gives rise to the question: Might emoji become the new lingua franca of the Internet?

A NEW FORM OF EXPRESSION?

Some might agree that emoji is becoming the new universal language of Marshall McLuhan’s “global village.” For example, the Oxford Dictionary declared the “face with tears of joy” emoji 😢 as its “Word of the Year” in 2015 (https://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2015). This is in recognition of the widespread global acceptance of the emoji as a popular means of expressing ideas and sentiment in an otherwise dry world of emotionless technocratic prose. The fact that a robust Unicode standard exists for emoji further reinforces a sense of universality. The need for maximum compression in Tweets, social media, text messages, and other digital media strongly encourages an economy of characters needed to express basic concepts. Whereas alphabets provide a finite set of characters to express any idea, many characters must be combined to do so. Emoji, at 144 pixels and 18 bytes, easily replace costly words with far greater economy.

Advertisers, quick to pick up on trends, regularly target Internet users with hip emoji messages. The level of monetization even extends to the service economy where employees are encouraged to quite literally present a smiley face to their clients, much less to cope emotionally in an otherwise insensitive world. Emoji appear to have “staying power” as an enduring visual code.

Below is a list of a number of useful emoji-related websites.

- Unicode Emoji Standard V 11.0: https://unicode.org/emoji/charts/full-emoji-list.html
- Real-Time Twitter Emoji Usage Tracker: http://emojitracker.com
Intended meanings of many emoji, however, can too easily be misconstrued. While the Unicode standard defines “core emoji,” many more (less well-defined) emoji continue to emerge daily worldwide. Soon a set of scientific emoji are poised to appear. This leads to a bit of a tower of Babel situation, as emoji are often culturally or contextually dependent. In fact, cultures with varying economic descriptors as defined by the Hofstede Culture Index are liable to use emoji differently to describe their particular relationship to the world. For example, people from countries with high uncertainty-avoidance scores tend to disfavor emoji that express positive emotion. Moreover, the same emoji might carry different meanings as determined by the culture where it is being used.

While the Unicode standard for emoji tends to reinforce meaning, there are at least 17 different proprietary platform-based fonts in place that significantly render the same Unicode emoji differently. The Unicode site (https://unicode.org/emoji/charts/full-emoji-list.html) shows 11 different platform-based renderings of standard emoji. Thus, a given standard Unicode emoji can appear differently on iOS than it does on an Android device. This leads to statistically different interpretations of both sentiment and meaning when specific standard emoji codes cross platforms. Nonetheless, variation in interpretation also occurs within the same platform, although to a lesser degree.

The rather generalized lack of commonality in emoji interpretation suggests that emoji are actually less than a universal form of expression. As noted, cultural influences, context, and symbolic variation can potentially compromise intended meaning. Worse, it would appear that emoji are less than a complete form of expression.

Standardized emoji codes do not really exist for personal pronouns or most intransitive verbs. This limits the expressiveness of the language, while simultaneously opening the door for creativity in usage among various user cliques. It is the case, however, that volunteers using Amazon’s Mechanical Turk encoded the entire text of Melville’s *Moby Dick* into a book entitled *Emoji Dick. Moby Dick’s* iconic first sentence, “Call me Ishmael,” was emoji encoded as follows:

```
Call me Ishmael.
```

*Figure 2. The first sentence of Moby Dick in emoji.*

While clearly a period novel, the use of a telephone (a nonexistent item in the time of the novel) induces a form of contextual irony. Likewise, *Alice in Wonderland*, a rebus-friendly text by the intent of author Lewis Carroll, has also been translated fully into emoji. In both cases, however, the level of effort necessary to successfully navigate these annotated texts exceeds the ability of most readers. Emoji datasets, while highly creative, become highly subjective, induce repetition, and become exceedingly difficult to contextualize. In other cases, multiple emoji must be creatively combined to suggest common items. For example, “sweetheart” might be written as a piece of candy next to a heart, hardly a literal translation.
Ultimately, emoji are technically oriented. As such, they are driven by advancing technology. Thus, as natural language processing (NLP) and artificial intelligence (AI) join forces to reinforce the effectiveness of vocal interaction, emoji might give way to vocalized inflections. Moreover, the number of bot-generated emoji could potentially overpower human users, much like spam often overwhelms the inbox. Both trends could signal a setback for emoji advocates.

The notion of emoji as an emergent universal language seems to be limited at best. The use of emoji as a hybrid form of expression to augment regular text, however, appears to be a strong and growing possibility in a world that increasingly demands symbolic economy and some level of personalization. Together with otherwise impersonal texts, selective use of emoji sets the tone for satisfying communication. Emoji tend to defuse what otherwise might be considered offensive messages with a friendly salutation, closing, or strategically placed emoji intended to add a more conciliatory tone.

**EMOJI IN A NETWORK AGE**

Emoji represent a network phenomenon. An analysis of an early August 2018 snapshot of the frequency of emoji usage on Twitter using the website http://emojitracker.com reveals a clear power curve relationship. Figure 3 shows this plot in the form of a vertical bar graph.

![Figure 3. Distribution of 846 popular emoji on Twitter in early August 2018.](image)

In this figure, the emoji occupying the top position was the familiar “face with tears of joy.” This emoji was invoked 2,145,510,490 times. The emoji at the last-used position, number 846, was called only 132,848 times. It was an emoji for uppercase Latin letters. The top 10 emoji were: face with tears of joy, a single heart, the recycling symbol, face with hearts for eyes, a slimmer single heart, a sad crying face, a simple happy face, a face with a furrowed brow and frown, a double heart, and a kissing face (see Figure 4).

![Figure 4. Top 10 emoji on Twitter in August 2018.](image)
It is interesting to note that the majority of the popular emoji are positive in nature, which is in keeping with most research on the use of emoji. Other research shows that applied network science techniques outperform state-of-the-art methods, including NLP for sentiment analysis.7

As noted, printing introduced a prevalent linear relationship that helped usher in an industrial age, enhancing the world’s economy. The advent of mass media, especially the Internet, awoke other sensitivities. The rise of the emoji as a popular means of visual expression suggests a return to age-honored visual space. Moreover, despite distinct cultural differences in usage, the world-wide emoji acceptance is itself significant. It represents a broad-based trend toward the reality of networked global sharing. Steeped in older linear technology models, many people fail to appreciate or perhaps even fear such openness. To some, emoji represent nothing short of a tragic fallback to primitive behaviors. Further social research along these attitudinal lines might better help further delineate growing protectionist movements in many nations.

As industrialization engaged, literacy-focused education became indispensable. Now formal education increasingly seeks creative online outlets, and traditional literacy-based instruction seems somehow outdated. Yet computer literacy continually gains credence. Importantly, the growing cost of formal higher education leaves many indebted well beyond any entry-level thresholds. Perhaps it is time to acknowledge the shift from book-borne portable personal knowledge to online networked general knowledge. Such a shift likely has a profound effect on future educational strategies. Here, new forms of digital literacy become prerequisite for future opportunity. The increased and sustained use of emoji might suggest new innovative research initiatives to help identify new educational vectors, perhaps even extending to mathematics.8

Finally, visualization is endemic. For example, most nations regulate driving behavior by varying shapes and color cues. Emoji only represents one form of the resurgence of visualization in the digital world. As a case in point, augmented reality and virtual reality are opening new perceptual doors. Graphical representation of data is also increasingly pressing. Networks of all types frequently involve large sparse matrices. The ability to visualize these diverse datasets becomes an increasingly critical skill. Conceptualizing and constructing such graphs require new mathematical insights and new means of depicting their hidden realities accurately and convincingly. More importantly, the ability to evaluate and interpret such visual representations on their merit is equally important for an informed citizenry.

To this end, the ability to acquire visual literacy, including the use of emoji, becomes an increasingly important skill—not only for dedicated data scientists, but across virtually all the increasingly entwined domains of human knowledge.

REFERENCES


ABOUT THE AUTHOR

George Hurlburt is chief scientist at STEMCorp, a nonprofit that works to further economic development via adoption of network science and to advance autonomous technologies as useful tools for human use. He is engaged in dynamic, graph-based IoT architecture. Hurlburt is on the editorial board of *IT Professional* and is a member of the board of governors of the Southern Maryland Higher Education Center. Contact him at ghurlburt@change-index.com.

This article originally appeared in *IT Professional*, vol. 20, no. 5, 2018.

IEEE Annals of the History of Computing

From the analytical engine to the supercomputer, from Pascal to von Neumann, from punched cards to CD-ROMs—IEEE Annals of the History of Computing covers the breadth of computer history. The quarterly publication is an active center for the collection and dissemination of information on historical projects and organizations, oral history activities, and international conferences.

www.computer.org/annals
In the era of “big data” there is an unprecedented increase in the amount of data collected in data warehouses. Extracting meaning and knowledge from these data is crucial for governments and businesses to support their strategic and tactical decision making. Furthermore, artificial intelligence (AI) and machine learning (ML) makes it possible for machines, processing large amounts of such data, to learn and execute tasks never before accomplished. Advances in big data-related technologies are increasing rapidly. For example, virtual assistants, smart cars, and smart home devices in the emerging Internet of Things world, can, we think, make our lives easier. But despite perceived benefits of these technologies/methodologies, there are many challenges ahead. What will be the social, cultural, and economic challenges arising from these developments? What are the technical issue related, for example, to the privacy and security of data used by AI/ML systems? How might humans interact with, rely on, or even trust AI predictions or decisions emanating from these technologies? How can we prevent such data-driven intelligence from being used to make malicious decisions?

Authors are invited to submit original, unpublished research work, as well as industrial practice reports. Simultaneous submission to other publication venues is not permitted. All submissions must adhere to IEEE Publishing Policies, and all will be vetted through the IEEE CrossCheck portal. For full CFP and conference information, please visit the conference website at WWW.COMPSAC.ORG

IMPORTANT DATES
April 7, 2019: Paper notifications
April 15, 2019: Workshop papers due
May 1, 2019: Workshop paper notifications
May 17, 2019 – Camera ready submissions and advance author registration due

ORGANIZING COMMITTEE
General Chairs: Jean-Luc Gaudiot, University of California, Irvine, USA; Vladimir Getov, University of Westminster, UK
Program Chairs in Chief: Morris Chang, University of South Florida, USA; Stelvio Cimato, University of Milan, Italy; Nariyoshi Yamai, Tokyo University of Agriculture & Technology, Japan
Workshop Chairs: Hong Va Leong, Hong Kong Polytechnic University, Hong Kong; Yuuichi Teranishi, National Institute of Information and Communications Technology, Japan; Ji-Jiang Yang, Tsinghua University, China
Local Organizing Committee Chair: Praveen Madiraju, Marquette University, USA
Standing Committee Chair: Sorel Reisman, California State University, USA
Standing Committee Vice Chair: Sheikh Iqbal Ahamed, Marquette University, USA
The practice of using disinformation and misinformation to promote parochial agendas isn’t new. Both have been used by tyrants, demagogues, dictators, authoritarians, and manipulators of every stripe for millennia. One thing that’s new to our generation is the digital twist of Internet trolling. The effectiveness and increasing use of this tactic, highlighted in the 2016 US presidential election, justifies increased attention. An earlier Computer column encouraged such attention, and we elaborate here.

Disinformation and misinformation both involve the distribution of false information, but with differing objectives. Disinformation involves the intentional planting of false information to conceal truth or deceive the audience, especially by state actors, whereas misinformation is more generic and relaxed regarding intention, concealment, and source. For our purposes, we intend the definition of disinformation to include not just governments but also political groups, ideological movements, and other social entities. Disinformation is more pernicious, being necessarily both intentional and deceptive in its pursuit of social engineering goals. Although some trolling might be without willful deception (as in the case of mistaken “true believers”), disinformation is the more natural ally of trolling and is thus our focus.

The topic of disinformation is both complex and varied: it’s complex owing to its convoluted methods; it’s varied because of its different practitioners and contexts. It can be used to enlist support, confuse, de-legitimize, defame, intimidate, confound, escape detection or blame, avoid prosecution, and on and on. The public relations strategist uses disinformation in different ways than the tyrant owing to the latter’s assumed greater imperviousness to punishment or retribution. Similarly, the ideologue’s use of disinformation is different from that of the corrupt politician. Disinformation techniques and content vary with the purpose, targeted demographic, medium, and social networking platform.

The Online Trolling Ecosystem

Hal Berghel, University of Nevada, Las Vegas
Daniel Berleant, University of Arkansas at Little Rock

As trolling becomes inseparable from modern social media, a renewed effort is needed to unmask and abate the risks of this reality. A proposed taxonomy offers useful clarification.
These issues apply to trolling as well. Consequently, we’ve developed a partial taxonomy to better characterize trolling’s many manifestations. This is an appropriate time for a taxonomy, for trolling is mature enough now to reveal interesting patterns and suggest future trends and defenses.

ROOTS AND MISSING LINKS

Trolling is confirmation, in a sense, of a fundamental flaw in the notional roots of the modern Internet-enabled Web. Those roots are typified by, for example, Paul Otlet’s Mundaneum system, implemented in 1910 to collect and categorize all of the world’s important knowledge (www.mundaneum.org/en); H.G. Wells’s notion of a World Brain, outlined in a 1938 collection of essays and addresses with that title; and Vannevar Bush’s Memex system, described in his influential 1945 article “As We May Think.”

Bush envisioned a collective memory system that would advance a knowledge explosion by serving up the corpus to anyone on demand through associative indexing and browser history-like “paths” not unlike the use of hypertext to organize the Web. As was customary in the early information age, Bush was driven by the simultaneous desire for ease of information access and avoidance of information overload. He wasn’t concerned about data reliability and source authentication.

As it turns out, this overly simplistic and naive view of the information access challenge has been perpetuated ever since on the Web. To wit, subsequent work on metadata standards, including the Dublin Core elements (http://dublincore.org/documents/dces; https://tools.ietf.org/html/rfc5013), completely ignore any measure of authenticity and reliability. The closest metadata elements would include oblique terms such as “provenance,” “conforms to,” and “is referenced by.” This deficiency has been carried forward in such subsequent document type definitions as the Open Source Metadata Framework and the Resource Description Framework. To overcome this deficiency, more user control is needed—perhaps a user-driven metadata insertion tool for elements like “suspect,” “disproved,” and “content warning,” or some sort of Bayesian trigger to deal with today’s fake news and alt-facts. Otherwise, the 21st century’s spin on Bush’s vision might progressively become “As We May Deceive.”

The study of disinformation, from an information-theoretic point of view, has thus far regrettably been at best occasional and informal. We have in mind, for example, contributions by David Martin and H. Michael Sweeney on disinformation and traits of disinformationists. While informative, especially with respect to the current political landscape, these works are largely anecdotal, lack examples, and aren’t directly related to trolling. Spy the Lie provides a practical guide, with examples, for detecting deception, including an analysis of behavioral cues that might betray the act. A rough equivalent for social media deception is sorely needed. Alas, self-published contributions on the Web, and those from the popular press, fail to do justice to the full impact of disinformation generally and trolling in particular.

TROLLING AS AN IDEOLOGICAL WEAPON

Online trolling is readily weaponized—it fits comfortably within pathocracy and kakistocracy as an effective tool of online manipulation, obfuscation, and deceit. Communication and the Russian government’s embrace of trolling. That said, the White House’s proneness to misinformation and even outright disinformation is a symptom of a more general social problem—namely, political emotionalism, in which facts are too often considered less of a foundation and more of a hindrance. That trend manifests itself in a tolerance of falsehoods under the guise of alt-facts, the inability to distinguish confirmable statements from beliefs and opinions, and an unreflective commitment to ideology-based and simplistic slogans, catch phrases, sound bites, formulas, and beliefs. Social scientists have developed theories of social dominance, authoritarianism, and instability that explain some these characteristics in terms of group behavior, economics, and social hierarchy.
WHY DISINFORMATION? WHY TROLLING?
Disinformation generally and trolling specifically are expedient ways to manipulate public opinion. Authoritarians of all generations understood that sound and reasoned argument isn’t sufficient to exercise control over others. Something more powerful but short of force is needed. Such machinations, to be effective, must be carefully engineered and targeted, an objective often unachievable through reasoned public debate. If politicians were to rely on logical debate, free of manipulative rhetorical devices, public consensus might be influenced by the merits of the arguments themselves when interests, often authoritarian or domineering, wish to avoid this.

Disinformation and trolling are expedient ways to manipulate public opinion. They can polarize issues to exploit a human bias toward binary choices.

Carefully crafted disinformation campaigns and trolling efforts can be instrumental in achieving the desired effect. They can artificially polarize issues to exploit a human bias toward binary choices—seeing the world in black and white, big and small, rich and poor. This is related to what Hans Rosling calls the gap instinct. Its appeal must follow in part from the cognitive simplicity of binary distinctions, much as we experience with true/false questions on exams. Other things being equal, cognitive effort is lower on true/false than multiple-choice questions because there’s less to think about.

Disinformationists and trolls seek to create a sense of extremes where the extreme they tout is cast in a more appealing way than the alternative. In order to force the information consumer to the desired extreme, they use lies, prevarications, untruths, alt-facts, unlikely theories, distortions, ad hominem attacks, and other rhetorical devices as part of a Machiavellian propaganda or “messaging” campaign to create the desired artificial duality in lieu of the more nuanced and reality-based presentation that would result from clear-headed analysis. Modern online disinformation and trolling campaigns functionally resemble phishing attacks in combining a modest amount of computing and networking skill to cloak the real goal and lure the target using perception management (manipulating the public into thinking they perceive something they don’t, or vice versa) and social engineering (motivating the public to do something they otherwise wouldn’t have done).

In his book Factfulness, Rosling describes how evolutionary traits like hard-wired fast-response brains produce simplistic world views that discourage adequate reflection and deliberation for decision making. He identifies 10 evolutionary “instincts” that no longer serve humanity well in separating truth from predatory fiction. Such instincts should be critically discussed as part of college-level general education, if not in high school. Primary education should provide practical skill in BS detection, right along with the 3 Rs. Call it the 4th R: reality checking.

A TAXONOMY OF TROLLING
Online trolling has matured to the point that we can discern some evolutionary patterns and future directions. The value proposition is obvious from the 2016 US presidential election: low-cost, potentially high-impact voter manipulation through micro-targeting. Political scientists and others continue to study the degree to which trolling influenced the vote. UK-based Cambridge Analytica executive Mark Turnbull took credit for playing a key role in Donald Trump’s win and there’s now sufficient concern over the use of trolling by foreign governments to undermine US federal elections that, as part of the Mueller probe, the US Department of Justice indicted the Russian trolling factory, the Internet Research Agency, for 8 federal crimes as well as 13 Russians and 3 Russian companies for attempting to subvert the 2016 election.

One thing is certain: online trolling is here to stay. Even if federal legislation were passed to outlaw it, problems like reliable cyber-attribution—at least that which is admissible in court—will provide trolls many avenues to circumvent whatever laws might be enacted.

So what’s the future of online trolling and its containment? We offer the following informal taxonomy as a means to focus our response.

Provocation trolling. To elicit a particular response, such as hostility, from participants of an online forum. For example, in the “Reactions” section of a Yahoo! article about a 20-year-old Guatemalan woman shot dead in Texas by a US border agent, many top comments seemed intended to spark a flame rather than shed light. For example, the first comment was “Medal of Honor!!!” Similarly, in an online discussion, blaming liberals or conservatives for a tragic or controversial incident will likely cause some offended readers to lunge for the bait.

Social-engineering trolling. To incite participants to activities they normally wouldn’t have undertaken—convince readers to join an organization, send a donation, observe a boycott, vote for/against a candidate, and so on.

Grooming trolling. Sending messages intended to insinuate the sender into the mind of the recipient as a slippery slope to further persuasion. Radical organizations are notorious for
using this variant of social-engineering trolling to recruit members: ISIS was widely noted for “fishing” for new members on Twitter this way, and US extremist groups are frequently noted for using this tactic.

**Partisan trolling.** To use social media surreptitiously to achieve political ends. Here’s where the heavyweights really get involved. For example, trolling has been exposed as an important component of Russia’s “firehose of falsehood” (see below) propaganda strategy, especially in the recent US presidential race.20

**Firehose trolling.** High-volume, rapid, continuous trolling without concern for consistency. Apparently a favorite of Russia, it focuses not on promoting a particular position or viewpoint but on divisiveness for its own sake. For example, according to Charles Clover, Aleksandr Dugin’s book The Foundations of Geopolitics is influential at the highest levels of the Russian government and “assigned as a textbook at the General Staff Academy and other military universities in Russia.”21 (A good English translation of the entire book isn’t yet available.) Clover quotes Dugin as writing, “It is especially important to introduce geopolitical disorder into internal American activity, encouraging all kinds of separatism and ethnic, social and racial conflicts, actively supporting all dissident movements—extremist, racist, and sectarian groups, thus destabilizing internal political processes in the U.S.” Trolling is certainly well suited to this activity. And it can be tough to counter. Christopher Paul22 recommends against trying “to fight the firehose of falsehood with the squirt gun of truth,” but fails to provide fully satisfying alternatives.

**Ad hominem trolling.** Defaming or discrediting individuals or groups to delegitimize their positions without engaging them on their merits. The following snippet from an exchange on an email list exemplifies this.

ML: [Controversial claim] Anybody who claims otherwise is ignorant, uninformed, or lying.

A naive respondent might be whiplashed at this point because a counter-argument, reasoned or not, has already been pre-characterized as ignorant, uninformed, or a lie. The best response is probably to simply point out the rhetorical device used here, as respondent PD does next.

PD: Ooh—is this the choose-your-own-ad-hominem part of the show?

Yet even this response is hobbled because the discussion has now been diverted into a rhetorical cul-de-sac that saves ML from losing the argument.

**Problems like reliable cyber-attribution will provide trolls many avenues to circumvent whatever laws might be enacted against trolling.**

**Jam trolling.** Disrupting a discussion or communication channel with high message volume (the trolling equivalent to a DOS attack). Technologically, automated trollbots will make this an increasing problem.

**Sport trolling.** Trolling for the self-gratification of the troll (just for the fun of it).

**Snag trolling.** Evoking responses to satisfy curiosity. One of the less toxic varieties, this nevertheless tends to divert and obscure.

**Nuisance trolling.** Derailing the thread of an online forum (blog, chatroom, and so on) for no other reason than to irritate other participants. A variant of sport trolling.

**Diversion trolling.** An insidious tactic for blocking legitimate communication by diverting a thread in a direction that’s misleading, irrelevant, false, and so on. Thus, a discussion about rising crime rates could be diverted by citing a small community that hasn’t had a murder in 20 years, or a discussion about falling crime rates could be diverted by mentioning a recent crime.

**False-flag trolling.** Pretending to be of a group or hold an opinion that the troll actually opposes, and presenting a message intended to make that group or opinion look bad. This is one of the harder forms of trolling to detect, because the writer could in theory really have the opinion claimed but not realize how his obnoxiousness is creating the opposite of the desired effect. For example, a type of robocall used in political campaigns pretends

---

www.computer.org/computingedge
providing the spirited practice the troll wants. The troll thus hones debate skills for uses like higher-stakes trolling later.

**Proxy trolling.** Using intermediary trolls to do the heavy lifting. De rigueur for large organizations, which hire people to do it.\(^{23}\) One application is astroturfing: promoting a position, product, person, and so on for which there’s little awareness or support by making it look like that entity is widely approved of. Websites and organizations set up by special interests but given names like “Citizens for X” are standard examples. Proxy trolling provides rich opportunities for all manner of resource-rich, unscrupulous actors.

**Faux-facts trolling.** Deliberate spreading of fake news, alt-facts, and other lies under the guise of truth. To fight with “Right on!” or “Thank you for saying what so many know but are afraid to say.” This boosts persuasiveness via a bandwagon effect.

**Chaff trolling.** Sending messages that are essentially content free and thus vacuous. For example, on social media platform Quora someone claimed that a relative assigned to help guard former president Obama said that the president was “... fake as [expletive deleted].” One might well question if this relative really existed, and if he did, whether the quote was accurate. Yet consider also the word “fake”: here it carries little if any information about its subject but is an effective insult for the many unsavvy readers.

**Wheat trolling.** High-quality trolling using content that’s hard or impossible to refute—for example, a cleverly doctored photo or text incorporating seemingly well-sourced “facts.” Some lies contain their own logical inconsistencies; others smell bad only to a domain expert.

**Insult trolling.** Insults spark responses that drain the target’s energy. They also make the target look bad and are demoralizing.

**PR trolling.** Making the troll or the views the troll is promoting look good rather than attacking others. For example, the troll could make a claim and unverifiably cite a brother-in-law “who was there.” But the most common example is to state approval of another text. It’s easy to upvote another troll’s message, or respond to a posting that’s indistinguishable from a human troll as a Turing trollbot—one that has passed the trolling equivalent of the Turing test. A computer-controlled chatbot passes the traditional Turing test if and only if the human tester cannot distinguish the chatbot from a human. Compared to a chatbot, a trollbot has a much easier time passing—the weaker constraints on trolling make it so. Sure, there are human trolls for whom sophisticated trolling is an unsavory art form that would be hard to imitate, but a Turing trollbot need only mimic the lowest-common-denominator human troll to masquerade as a real person.

The concept of the Turing trollbot is increasingly recognized.\(^ {24}\) The hardest technical aspect of primitive Turing trollbot design is sneaking through smart filters like CAPTCHA. In fact, such trollbots could soon emerge as easily downloaded freeware apps. But primitive Turing trollbots are just a start. As we were writing this article, IBM unveiled its Debater system,\(^ {25}\) which successfully took on a college debate champion. This is a much greater challenge than deploying successful trollbots, which can be ever so much more efficient and economical than a paid human.

With armies of well-nigh undetectable trollbots on the horizon, what’s one to do against this threat? One approach is to simply ignore outright all controversial social media comments—that might protect individual readers. Another approach is mass immunization. The simplest way to ensure public health is for enough people to reply to suspected troll messages by shining a light on them. “Are you a troll?” might serve not just as a comment but as a warning and reminder to readers who otherwise might have overlooked the possibility. But one way or another, society must
Research is also needed to investigate the potential for automatic trolling detection software. What kinds of trolling are undetectable? What kinds have already been detected, and who are their sponsors? We also need to educate the public. An increasingly necessary goal of primary education is training people to approach social media statements with suspicion, especially when it comes to bias and misinformation. The Internet—through social media and fake news outlets—has saddled us with the biases of those seeking to manipulate others through new forms of information corruption such as source displacement/concealment, decontextualization, and the like. Where the traditional measures of networks were in terms of value,\textsuperscript{26,27} a new and useful measure of networks is their potential for abuse.\textsuperscript{28}

**POLITICAL TROLLING**

In addition to the computer and networking context, online trolling must be understood in a geopolitical context\textsuperscript{29,30} especially with respect to its utility in international competition and rivalry. For example, a measurable amount of the identified external political trolling used to influence the outcome of the 2016 US election appears to have been either sponsored or inspired by Russia. China certainly has the capability for effective political trolling as well. As time passes, more countries will inevitably engage in it as a useful and cost-effective way to project influence. Free societies are the most susceptible to political trolling because in those countries mass opinion is a strong driver of national policy.

Moreover, polarization and partisanship have been increasing for decades.\textsuperscript{11,31–33} Trolling’s utility is related to the political divisiveness of the target society. As trolling and other ways of abusing social media and networks evolve, the current deficiencies in teaching disinformation tactics widely as an important civic skill will become more apparent. Our children, like all too many adults, lack the basic skills to look upon divisive, emotive communication critically. This is a severe educational shortcoming that promises to exact a considerable toll on democratic systems.

Society needs to understand why people troll. It seems to be one of many addictive behaviors mostly afflicting alienated young males and enabled by the anonymity and easy accessibility of the Internet, much like overindulging in online porn or videogames (https://www.quora.com/Whats-it-like-to-be-an-Internet-troll). But perhaps it’s not as important to understand the psychology underlying trolling as it is to avoid being manipulated by it. As Lee Edwin Coursey\textsuperscript{24} advises,

\textbf{The next time you see a hyperbolic social media post that confirms your worst fears about people of a particular race, gender, religion, or political affiliation, your first reaction should be, “nice try, Russian troll,” rather than “OMG I MUST REPOST THIS EVERYWHERE!!!” Learn to take a breath and pause before you immediately like, retweet, or share divisive messages from obscure sources. Be especially wary of emotional manipulation. Most importantly, fact check yourself before spreading information designed to foment outrage and factionalism. Remember that the phrase “Russian disinformation campaign” does not describe some outdated method from a bygone era, but instead represents an active, effective tool being used against you right now.}

College outcome of the 2016 US presidential election. This is where trolls and other social media manipulators see the real payoff. It’s for this reason that so much trolling content tends to be shocking, distressing, offensive, and the like—it’s designed to arouse the passions of the recipient while not lending itself easily to deliberation. The more independent fence-sitters can thus be stimulated to action or opinion without benefit of the reflection that would call into question the validity of the message or stimulate thoughtful evaluation. Fact checking, introspection, and analysis work against the interests of trolls. In this way, trolling is similar to a military campaign where the goal is action without debate.

We might take a lesson from Winn Schwartau’s Time-Based Security Model
in this regard. The model posits that a security system can be effective only when the time it takes to detect a security breach and mitigate against the threat is less than the time it takes for the security breach to achieve its objective. There's a parallel when it comes to mitigating against the effects of abusive social media. For it to be effective, the detection time must be near zero because the reaction time required to re-tweet, forward, and so on is negligible. The parallel with trolling is that the troll is focused on achieving quick results before second thoughts might be raised.

It’s worth adding that trolling’s ability to promote division can also be used to nurture social reform and is thus a doubled-edged sword for authoritarian and totalitarian states. For that reason, such states must carefully monitor and control trolling and related digital media manipulation tools within their borders. New though it is in the toolbox of Machiavellian kingpins and social misfits alike, the effectiveness of trolling ensures that it’ll continue to play an important role in future politics.

REFERENCES
in this regard. The model posits that Machiavellian kingpins and social divisions within their borders.

is thus a doubled-edged sword for authorities. It's worth adding that trolling's ability to promote division can also be used to nurture social reform and quick results before second thoughts might be raised. For it to be effective, the detection time must be.

A security system can be effective only if the detection time is short enough to prevent a security breach and mitigate against the Russian 'Firehose of Falsehood,' as RAND's Christopher Paul discusses.

This article originally appeared in Computer, vol. 51, no. 8, 2018.
CareerVis: Hierarchical Visualization of Career Pathway Data

We present our CareerVis system, an interactive visualization tool to aid career education for high school and freshman college students. In addition to its practical use, we believe our design approach has potential to inspire the design community to develop simple visualizations that convey complex information to novice users.

To help students prepare for success in college, career, and life, education stakeholders have a long-term commitment to developing curriculum that allows students to explore their college and career options as well as their aptitudes and employability (www.doe.in.gov/sites/default/files/standards/cte-family-and-consumer-sciences/cf-busfacs-preparingcc_7-11-14.pdf). Studies have shown that the right kinds of education can help people ease their transition into the job market. Students should be able to plan for college and career pathways that are suitable for their interests, abilities, and lifelong goals.

Within this context, the community needs efficient tools to help students better prepare for their careers after graduation. Based on a synthesized dataset from Purdue graduates’ job placement survey data and a national survey database (www.onetonline.org; www.mynextmove.org), we developed CareerVis, a visualization system aimed to help young students comprehend the broad range of educational and occupational paths as part of the college-career selection process.

In order to improve the efficiency of decisions made by students, parents, and other career education stakeholders, we need to anticipate the following frequently-asked questions:
1) Which major should I choose? Which occupation should I pursue?
2) What majors can help me pursue this occupation?
3) What occupations am I qualified for with this major?
4) What are the characteristics of these majors and occupations?

THE DATA AND MESSAGE

Our underlying dataset contained flow information with hierarchical structure and multidimensional characteristics, which could be decomposed into several typical data structures in information visualization design[^4]. Particularly, the dataset was composed of college majors, occupations, flows of students from majors to their first jobs, 12 numeric measurements for majors (e.g., GPA and SAT scores), and 12 measurements for occupations (e.g., salaries and future trends in globalization and automation). The dataset presented the following challenges on visualization:

1) three different types of data formed a more complex data structure;
2) the relatively large amount of data;
3) the data would be presented for the general public, who have little experience with reading and understanding visualization applications; and
4) the visualization would also be published on a relatively small screen (e.g., a tablet or smartphone).

The data is inherently hierarchical. Purdue University’s West Lafayette campus houses 145 departments within the 10 colleges: Agriculture, Education, Engineering, Health and Human Sciences, Liberal Arts, Management, Pharmacy, Science, Technology, and Veterinary Medicine. Students’ occupations were aggregated into 130 specific job positions and put into occupation groups by the standard occupational classification system ([www.bls.gov/soc](http://www.bls.gov/soc)). Moreover, the data contains proportions of the student body enrolled in majors or landed in a different job.

There are multiple pathways for students to pursue their ideal occupations from majors. For instance, among students who received accounting degrees, 70% secured accountant and auditor positions, while 17% worked as financial analysts, and 1%-3% worked in 8 other occupations. Vice versa, other than two majors (Accounting and General Management) under Management, about 10% of accounts come from majors under six different colleges which range from Agriculture to Science. There are about one thousand possible major-to-occupation pathways.

College majors and occupations have many numerical measurements that can offer students insight into the requirements of certain majors (e.g., GPA, SAT scores required for enrollment), important measurements for jobs (e.g., salary and work hours), characteristics of certain majors and occupations (e.g., percentage of Indiana students, percentage of domestic/international students, and diversity status of minorities), and future trends that could be affected by automation and globalization. These characteristics are complementary descriptions of majors and jobs necessary to guide students’ decision making. People may find certain jobs are better suited for family-oriented employees since their percentage of married workers is higher than other professions (e.g., engineering and construction). Some job opportunities may increase (or decrease) with the development of automation and globalization. Some characteristics have single percentage values (e.g., the percentage of the workforce from various ethnic groups), while others have percentile values of 10%, 25%, median, 75%, and 90%, such as salary, GPA, and SAT scores.

DESIGN EXPLORATION AND ITERATIONS

Our team conducted several iterative designs and involved users in the design processes. From more than ten design ideas [e.g., Figure 1(a)], we selected the one with the most simple and intuitive form to develop. Figure 1(b) shows our first formal design in horizontal layout. Characters are visualized as heat maps. Flows are presented in a hierarchical Sankey diagram. Our testing showed that the majority of users fully comprehended the hierarchies between the colleges and majors, as well as the...
occupational groups and occupations. However, the horizontal design presented several problems such as the interface layout generated unnecessary information overlap.

To better utilize screen space and represent characteristics, we developed another version. The main design idea of hierarchical flow remains but was rotated into a vertical presentation. To improve the design of side characteristics, we brainstormed many solutions [Figure 1(c)]. Ultimately, we selected a relatively intuitive option that featured the direct relationship between the line and value to show the characteristics of verbal, quantitative, and reasoning skills [Figure 1(d)]. The straight line would be highlighted with bright red when hovering on a particular major or occupation. Additionally, we presented other characteristics by bar graphs. Based on interviews with 64 participants, 90% of users were more satisfied with the vertical design. We further enhanced this design to address some remaining problems with the implementation of interactions, mainly the lack of a comparison function and the fact that once a user clicked on a college or an occupation group, the block expanded suddenly, which caused the user to lose visual momentum due to a sudden change of the visualization.

**VISUAL COMPONENTS**

Our resulting CareerVis user interface provided multiple views for data exploration [Figure (2)]. The central flow view displayed the hierarchical levels of colleges, majors, occupational groups, and occupations, as well as the one-to-one relationship between major and occupation. Each particular major or occupation featured a complete description. All relevant characteristics of majors and occupations were presented by scatter plots, box plots, and bar charts. In the top view, the system provided a guide and a search function. The system is developed using d3 (d3js.org).

Figure 2(a) shows the breadth of occupations chosen by students after graduating from college. In the central section, our team combined three essential visual components:

1) the rectangular blocks of the Purdue colleges and majors on the left side, where lengths of blocks represent the number of students that have graduated;
2) the similar visual element of the occupational groups and occupations on the right side, where lengths represented the number of students in the occupational group; and
3) the connection paths in the central region.
To improve the design of side characteristics, we brainstormed many solutions [Figure 1(c)]. The main design idea of hierarchical flow remains but was rotated into a vertical presentation. We further enhanced this design to address some remaining problems with the flow visualization part and keep it centered in the current window.

The paths that connected both sides show the percentages of students who graduated with a particular college major(s) and obtained a specific job position(s). The college major and occupation/occupational group were embodied in the two content-rich side bars. Clicking on a college expand the college into majors.

Characteristics plots [Figures 2(b), 2(c), 2(d), and 2(e)] provide users with detailed information about the colleges, majors, occupational groups, and occupations. For the verbal and quantitative scores, we use scatter plots, one for the college side and another for the occupational side. Dots in the plot represent majors or occupations. The color of a dot is based on the major/occupation’s requirement of verbal and quantitative skills. For characteristics like salary, SAT score, and GPA, we took advantage of their 10th, 25th, 50th, 75th, and 90th percentile data and use the box plot for their visualization. Users can not only compare the differences between majors and occupations, but they can also view the data distribution within the major or occupation.

The accordion6 [Figure 2(f)] was introduced to represent many characteristics of majors and occupations. With the accordion, the system can contain numerous characteristics’ charts within a relatively small space. Users can expand the characteristics they want and fold those in which they have no interest. An ordinary PC or laptop screen can hold four characteristics charts at the same time. The most important characteristics are opened by default on top of the list when users enter the system. If a user opens several charts that extended beyond the screen, they can scroll down the screen. The system automatically adjusts the position of the central flow visualization part and keep it centered in the current window.

Color is used to encode the two basic skills—quantitative skills and verbal skills—required by each occupation and major. If one major/occupation requires more quantitative skills, its color is
bluer. Yellow means the major/occupation requires balanced quantitative and verbal skills, whereas red means the major/occupation requires more verbal skills. The saturation of color represents the strength of skills these majors/occupations required. The quantitative/verbal Cartesian coordinate system is mapped into a hue (blue to yellow to red) saturation polar coordinate system. The radius (saturation) is computed by the distance of the quantitative/verbal (X/Y) to the origin (0, 0). The hue is computed by the angle of the quantitative/verbal point to the X-axis (verbal). Colors are featured consistently across all sections of the quantitative/verbal scatter plot, characteristic graphs, and center major/occupation flows.

INTERACTION AND ANIMATION

Brushing and Linking

Brushing and linking allow users to locate a major or occupation from the side charts and verify the corresponding values. The system’s brushing and linking occur when users hover over a bar in the center diagram, and the corresponding side chart elements are highlighted. When users hover over the College of Engineering in the center diagram, for example, the circle of majors in the corresponding scatter plot are highlighted by a thin black stroke and other circles fade [Figure 3(a)], and the bars of the College of Engineering are highlighted in every opened characteristic chart by a gray line. The values of each characteristic are shown at the top of each gray line [Figure 3(b)], which allows users to easily confirm the exact values. Brushing and linking are active within both levels of the chart, which means that when hovering over a college, occupation, or major with an open center chart, the corresponding elements are all be highlighted in the charts of the corresponding sides.

Figure 3. Brushing and linking of (a) scatter plot; (b) other characteristics charts.
Focus

A focus stage was added to the bars for when users hover over the parallel sets. When a bar is focused, the transparency of the paths linked to this bar increases, and the transparency of the other paths decreases to the extent that users can easily distinguish the focused paths from other paths and distributions. There are some differences in the interactions between the first and second hierarchies, depending on which one becomes focused, but the percentage of each path will be displayed.

Contexts and Details

The system constantly locates users and compares the current focus with related information. There are two categories of context maintenance. The first is used in the flow path of the Sankey diagram [Figure 4(a)] and the scatter plot [Figure 4(b)]; the second is applied to the bar set of the Sankey diagram [Figure 4(c)], box plots, and bar charts of the second-level display [Figure 4(d)]. For the flow path, as mentioned above, when a bar set is focused, the transparency of the connected paths increases, and the other paths become transparent. We maintain the paths as light background context elements. We use the same strategy of highlighting the selected circles and fade the others within the scatterplot. As for the two levels, we keep the first level elements displayed when the second level is opened. In the center Sankey diagram, the unselected first-level bars are shrunk and faded. In the box plots and bar charts, the unselected first-level bars are shrunk and moved to the left or right based on their position relative to the selected bar.

Figure 4. Zoom in for contextual details: (a) central flow of majors in one college to occupations in a group; (b) scatter plot highlighting focused majors and occupations; (c) context of other college and occupation groups; (d) detailed characteristics of occupations within the context of all other groups.
Animated Transition

The system’s animated transitions help the users understand the relationships between the two levels. The rich interaction causes the system to change its visual elements significantly. To maintain the cognitive coupling of the user with the system, we reinforce visual momentum by incorporating animated transitions in both the center diagram and the characteristic charts. For the center Sankey diagram, when a college bar is opened, the college bar gradually expands to a settled length, and the second college bar levels, which are the majors within this college, gradually expand and occupy the college bar area. At the same time, other college bars shrink to the same smaller size for readability [Figure 5(a)]. The occupation side applies the same rule. The transitional animations of the characteristic charts follow the same rule: the selected bar gradually expands and disappears, the corresponding second level bars gradually appear and expand, and the other first level bars shrink [Figure 5(b)]. All of these animations take 0.5 s, a value selected after user testing of different animation lengths.

![Figure 5](image_url)

Figure 5. Animation design context details of the (a) central flow and (b) two-side characteristics.

General Information Query With the Central Flow

Assume a user is a student from mechanical engineering (ME) and they want to find out what occupations they may have in the future. The user should first open the college of Engineering, mouse-over to ME to see connections to occupations several occupation groups [Figure 6(a)], then open the major occupation group of Engineers [Figure 6(b)] to see the most
frequent occupations from ME. The user can then close both sides to return to the initial flow view [Figure 6(c)].

Characteristics

Assuming a user is interested in the engineering major, they can compare the differences in verbal and quantitative skills required for the major. When they open the College of Engineering chart, several converging points, 0.65–0.9 verbal and 0.6–0.9 quantitative, are highlighted. When hovering on ME, the dots of 0.9 verbal and 0.9 quantitative are highlighted with a red stroke [Figure 7(a)]. Compared with ME, computer engineering requires on average 0.65 verbal and 0.8 quantitative [Figure 7(b)].

Moreover, when the user moves on to electrical engineering, they find that this major requires 0.75 verbal and 0.8 quantitative. They decide the best possible occupation for their major is Electrical Engineer based on these statistics [Figure 7(c)]. Furthermore, the user proceeds to the side box salary plot and notices that the median earnings for that occupation are $65k [Figure 8(a)]. They learn the average hours of work per week by an electrical engineer is 43 [Figure 8(b)], and the occupation’s value of automation sensitivity is 10 [Figure 8(c)], which means it is very sensitive to the development of automation. They can open up different characteristic tabs to develop their overall knowledge of the pathway. Instead of clicking on a college/occupation group in the center flow bar to see all detailed majors and occupations, the user can click on a box or a bar in the characteristic graphs to open up the college/occupation in the center graph.

USER FEEDBACK

In our usability testing, we designed quantitative and qualitative questions to compare participants’ understandings and expectations toward the relationship between college majors and occupations. The number of the valid participants was 68, all of whom were first-year students who came from Purdue University. We recorded their mouse activities during the experiments and only nominated the participants who spent more than one minute actively in our system, while the average spending time was more than 30 min. As the result, the overall compatibility
of participants’ selected majors and occupations increased from 79% to 91% after using our system. Additionally, 13 changed their primary major selection, and 30 changed their primary occupation selection. They further provided feedback about how they evaluated this application by ranking from best to worst various aspects of this application: attractiveness, clearness, colorfulness, helpfulness, difficulties, effectiveness, and efficiency. Their responses indicated that this application received the highest scores in being helpful, colorful, and attractive, with difficulty and efficiency being areas for improvement.

In 2017, the Indiana Department of Education proposed a set of learning objectives for the curriculum to prepare students for college and careers (www.doe.in.gov/sites/default/files/standards/cf-bus-facs-pcc-01-2016.pdf). Purdue curriculum specialists examined our CareerVis tool and found that the tool could help achieve these learning objectives and that the tool would be useful to integrate into the course curricula to improve college and career teaching and learning.

CONCLUSION

Based on recent years’ job placement data from Purdue students, we designed and developed our visualization system, CareerVis, to represent the vast choices of education and pathways to the occupation. Using innovative yet simple visual forms, we believe we solved two big challenges in the visualization. First, our system is able to deal effectively with hundreds of majors and occupations, and a thousand possible pathways. Second, our system has proven to be easily comprehensible by a general audience, such as high school students, parents, and educators. Collaborating with experts in education, we are now working toward integrating this system with curriculum to address the Indiana Department of Education’s high school academic standards for Preparing for College and Careers.

ACKNOWLEDGMENTS

This work was supported by a grant to Purdue University from the Lilly Endowment Inc. The grant is part of the Lilly Endowment’s “Round III: Initiative to Promote Opportunities through Educational Collaborations.” Access the CareerVis visualization system at: va.tech.purdue.edu/careerVis.

REFERENCES

ABOUT THE AUTHORS

Mingran Li is currently working toward the Ph.D. degree with the research focus on information visualization design at Purdue University, West Lafayette, IN, USA. Contact her at li1940@purdue.edu.

Wenjie Wu is a User Experience Designer. She received the Master of Science degree from Purdue University, West Lafayette, IN, USA. Contact her at wenjie.jessie.wu@gmail.com.

Junhan Zhao is currently working toward the Ph.D. degree at Polytechnic Institute, Purdue University, West Lafayette, IN, USA. Contact him at zhao835@purdue.edu.

Keyuan Zhou is a Research Assistant with the Polytechnic Institute, Purdue University, West Lafayette, IN, USA. Contact him at zhou851@purdue.edu.

David Perkis is Director of the Purdue Center for Economics, Purdue University, West Lafayette, IN, USA. Contact him at perkis@purdue.edu.

Timothy N. Bond is an Assistant Professor with the Economics Krannert School of Management, Purdue University, West Lafayette, IN, USA. Contact him at tnbond@purdue.edu.

Kevin Mumford is Director of the Purdue University Research Center in Economics (PURCE), West Lafayette, IN, USA. Contact him at mumford@purdue.edu.

David Hummels is Dean and Professor of economics with Purdue University, West Lafayette, IN, USA. Contact him at hummelsd@purdue.edu.

Yingjie Victor Chen is an Associate Professor with the Department of Computer Graphics Technology, Purdue University, West Lafayette, IN, USA. Contact him at victorchen@purdue.edu.

Contact department editor Mike Potel at potel@wildcrest.com.

IEEE Computer Society conferences are valuable forums for learning on broad and dynamically shifting topics from within the computing profession. With over 200 conferences featuring leading experts and thought leaders, we have an event that is right for you.

<table>
<thead>
<tr>
<th>Find a region:</th>
<th>Africa</th>
<th>Asia</th>
<th>Australia</th>
<th>Europe</th>
<th>North America</th>
<th>South America</th>
</tr>
</thead>
</table>

**JUNE**

2 June
- ICESS (IEEE Int’l Conf. on Embedded Software and Systems)
- JCDL (ACM/IEEE Joint Conf. on Digital Libraries)

3 June
- AIKE (IEEE Second Int’l Conf. on Artificial Intelligence and Knowledge Eng.)

5 June
- CBMS (IEEE 32nd Int’l Symposium on Computer-Based Medical Systems)

9 June
- WoWMoM (IEEE 20th Int’l Symposium on “A World of Wireless, Mobile and Multimedia Networks”)

10 June
- ARITH (IEEE 26th Symposium on Computer Arithmetic)
- ICHI (IEEE Int’l Conf. on Healthcare Informatics)
- MDM (20th IEEE Int’l Conf. on Mobile Data Management)

12 June
- ICIS (IEEE/ACIS 18th Int’l Conf. on Computer and Information Science)
- SMARTCOMP (IEEE Int’l Conf. on Smart Computing)
- WETICE (IEEE 28th Int’l Conf. on Enabling Technologies: Infrastructure for Collaborative Enterprises)

15 June
- CVPR (IEEE/CVF Conf. on Computer Vision and Pattern Recognition)

16 June
- SASO (IEEE 13th Int’l Conf. on Self-Adaptive and Self-Organizing Systems)

17 June
- EuroS&P (IEEE European Symposium on Security and Privacy)

21 June
- CSCloud (6th IEEE Int’l Conf. on Cyber Security and Cloud Computing)
- EdgeCom (5th IEEE Int’l Conf. on Edge Computing and Scalable Cloud)

22 June
- ISCA (ACM/IEEE 46th Annual Int’l Symposium on Computer Architecture)

24 June
- DSN (2019 49th Annual IEEE/IFIP Int’l Conf. on Dependable Systems and Networks)
- ECMSM (IEEE Int’l Workshop of Electronics, Control, Measurement, Signals and their application to Mechatronics)
- IC2E (IEEE Int’l Conf. on Cloud Eng.)
- LICS (2019 34th Annual ACM/IEEE
Symposium on Logic in Computer Science)

25 June
• CSF (IEEE 32nd Computer Security Foundations Symposium)

JULY
8 July
• ICME (IEEE Int’l Conf. on Multimedia and Expo) ▲
• ICMEW (IEEE Int’l Conf. on Multimedia & Expo Workshops) ▲
• SERVICES (IEEE World Congress on Services)
• SNPD (20th IEEE/ACIS Int’l Conf. on Software Eng., Artificial Intelligence, Networking and Parallel/Distributed Computing) ▲

15 July
• ASAP (IEEE 30th Int’l Conf. on Application-specific Systems, Architectures and Processors)
• CBI (IEEE 21st Int’l Conf. on Business Informatics) ▲
• COMPSAC (IEEE 43rd Annual Computer Software and Applications Conf.)
• ICALT (19th IEEE Int’l Conf. on Advanced Learning Technologies) ★
• ISVLSI (IEEE Computer Society Annual Symposium on VLSI) ▲

23 July
• ICCI*CC (IEEE 18th Int’l Conf. on Cognitive Informatics & Cognitive Computing)

30 July
• IRI (IEEE 20th Int’l Conf. on Information Reuse and Integration for Data Science)
• SMC-IT (IEEE Int’l Conf. on Space Mission Challenges for Information Technology) ▲

AUGUST
1 August
• CSE (IEEE Int’l Conf. on Computational Science and Eng.)
• EUC (IEEE Int’l Conf. on Embedded and Ubiquitous Computing)

5 August
• TrustCom (IEEE Int’l Conf. on Trust, Security and Privacy in Computing and Communications)

8 August
• 2019 Cloud Summit

9 August
• SmartIoT (IEEE Int’l Conf. on Smart Internet of Things) ▲

10 August
• HPCC (IEEE 21st Int’l Conf. on High Performance Computing and Communications)
• SmartCity (IEEE 17th Int’l Conf. on Smart City) ▲
• DSS (IEEE 5th Int’l Conf. on Data Science and Systems) ▲

15 August
• NAS (IEEE Int’l Conf. on Networking, Architecture and Storage) ▲

18 August
• HCS (IEEE Hot Chips 31 Symposium) ▲
• RTCSA (IEEE 25th Int’l Conf. on Embedded and Real-Time Computing Systems and Applications) ▲

27 August
• ASONAM (IEEE/ACM Int’l Conf. on Advances in Social Networks Analysis and Mining)

SEPTEMBER
13 September
• EWDTS (IEEE East-West Design & Test Symposium)

19 September
• AVSS (16th IEEE Int’l Conf. on Advanced Video and Signal Based Surveillance)
• ESEM (ACM/IEEE Int’l Symposium on Empirical Software Eng. and Measurement) ★ ◆
Mind your business.

Preserve your good name.

IEEE Member Professional Liability Insurance Plan.
Protect what you’ve worked so hard to achieve.

To learn more*

call 1-800-375-0775 or visit IEEEinsurance.com/IEEEPL

Canadian IEEE members can visit www.ieeeinsurance.com/canadapl for more information about the insurance program brokered by Marsh Canada Limited and underwritten by Certain Underwriters at Lloyd’s of London.

*The IEEE Member Professional Liability Insurance Program with the Choice Platform is available to active IEEE members who reside in the U.S. IEEE members in Canada (excluding Quebec) have access to the IEEE Member Professional Liability Insurance Plan through Marsh Canada Limited. Coverage options may vary or may not be available in all states. Not all plan features will be available under all carriers or plan options.

Program Administered by Mercer Health & Benefits Administration LLC
In CA d/b/a Mercer Health & Benefits Insurance Services LLC
AR Insurance License #100102691 | CA Insurance License #0G39709

85777 (5/19) Copyright 2019 Mercer LLC. All rights reserved.