Preface

China Research Laboratory

The IBM China Research Laboratory in Beijing conducts research in a wide range of areas, including business process integration and management, information and knowledge management, digital media and embedded systems, future client systems and user interactions, pervasive computing and on-demand infrastructure. In this journal issue, we present five papers from three of these areas. These papers cover a wide scope ranging from the exploration of future device scenarios which address projected market needs and positions IBM for the next generation of volume wireless messaging, to the pursuit of excellence in services through actual customer engagements. These topics are not only interesting in themselves but are also particularly relevant to emerging markets.

In the area of pervasive computing, we look at the infrastructure capabilities needed to support various pervasive computing and wireless applications. Wireless and mobile devices can provide both device identity and location information through various short-range and long-range location positioning methods. Location-based or location-aware services have been widely proposed and discussed. In the paper by Chen et al., a detailed description is given for an information infrastructure which is needed to support location-aware services. This location operating reference model can manage a vast amount of location information, provide modeling capability, and support querying and notification. Critical components of this infrastructure would include a location server, a mobile object database, and a spatial subscription and publication facility. A number of technologies can be used for determination of indoor locations, depending on the application scenarios, location precision required, and cost implications. Among several options, Wireless LAN (WLAN) is by far the most widely deployed infrastructure. The paper by Xiang et al. suggests a novel way of determining the position of a WLAN-enabled device down to a precision of two meters, while the WLAN access range can be tens of meters or more. Such solutions will be of interest to retailers who would like to push information to shoppers at specific locations.

For much wider ranges of communication, the wireless communication network is expanding beyond the current second-generation (2G) technology. Short Messaging Services (SMS), extremely popular in Asia and Europe, will be extended into Multimedia Messaging Services (MMS). The paper by Shen et al. proposes a scheme for adding interactive capabilities to MMS by embedding eXtensible HyperText Markup Language (XHTML) into the MMS presentation language to support good user experience and workflow. This provides an alternative mechanism for extended applications with asynchronous messaging between clients and servers as well as among peers.

In the area of future devices, we believe that there is no single device which can provide all of the functions, convenience (e.g., various form factors), and cost containment in the near future, while each device continues to become more powerful through embedded communication capabilities and software. This provides room for virtual devices which could be "dynamically" assembled by various intelligent information appliances to augment the total function and user experience. The paper by Fu et al. illustrates how such virtual devices can be created.

We have been looking for a novel method which can significantly enhance service capabilities for transforming business through information technology (IT). The paper by Zhu et al. illustrates in detail a model-driven business process integration approach which can significantly bridge the gap between business initiatives and IT implementation. In addition, the authors describe the insights developed through a customer engagement. This paper also provides a role model for describing the services-oriented research which will be a new frontier for researchers.

India Research Laboratory

The IBM India Research Laboratory in New Delhi pursues significant research initiatives in electronic commerce, information management, software lifecycle, autonomic computing, technologies for human–computer interaction, and solutions and services. Selected papers in this issue include one set of papers describing the uses of machine learning and information retrieval techniques for applications such as marketing research, checking consistency of Web sites, and analyzing biological documents. A second set of papers deals with issues that are of significant importance to information technology access and use in India.

In the first category, the paper by Agrawal, Basak, et al. describes a novel method for conducting marketing research online. Traditional marketing research, conducted through telephone interviews, by intercepting physical shoppers in retail stores or malls, and through focus groups, is expensive, slow, and prone to inaccuracies. There are an increasing number of online shoppers which can be used to gather market intelligence rapidly and cost-effectively. The authors discuss techniques such as active learning, matched control and experimental groups, and implicit and explicit experiments used to gather market intelligence online. These techniques can provide substantial value to online commerce Web sites.

Another important issue that must be dealt with by online commerce sites and Web sites in general is the

reliability, timeliness, and correctness of information on the site. However, typical commerce sites contain a large number of pages and publish information from multiple back-end data sources. Given the size of the site and the frequency with which the data sources are updated, it becomes essential to have an automated means of checking for correctness and consistency of data. The paper by Agrawal, Ananthanarayanan, et al. describes a method for monitoring Web sites to ensure consistency of content. The tool described in this paper allows users to specify items of interest to be tracked, monitors these items on the Web pages, and reports any changes observed. It is an end-to-end solution comprising a highly configurable crawler, a miner, and a reporting system.

In the biomedical field there exists a large volume of textual material in the form of scientific publications, clinical reports, patent applications, and so forth. The unstructured nature of such publications makes it difficult to utilize data mining or automated knowledge discovery techniques to extract information from these sources efficiently. In order to analyze these documents by machine, it is essential to annotate the documents appropriately. The paper by Mukherjea et al. describes a novel method for identifying and classifying biological terms in a document. Previously known terms are dealt with using a dictionary, and a rule engine is used for discovering new terms. The system also provides methods for disambiguating biological terms that belong to multiple semantic classes.

An important aspect of information access in regions such as India is the ability to provide sufficient support for access in regional languages. One of the major languages used in India is Hindi, which does not lend itself well to keyboard input. A speech recognition system then becomes a natural choice for accessing a computer. The paper by Kumar et al. describes techniques for building a large-vocabulary speech recognition system for Hindi. This paper describes a technique for bootstrapping an initial acoustic model by using trained models in another language. The paper also addresses the amount of time that is required to generate a dictionary of word pronunciations in the new language. This is tackled through a unique combination of pronunciation rules acquired from an expert linguist and statistical learning from training data.

Governments in many regions including India are increasingly leveraging the Internet and Web access to improve their operations and enable citizens to access information and complete government processes through direct electronic channels. In many cases these solutions are being developed independently of one another, with little sharing of resources or interlocking of processes between them. Mittal et al. present a framework for building such eGovernance solutions that emphasizes reusability and the use of open platforms. The solution framework described in this paper simplifies the task of developing, deploying, and managing complex eGovernance solutions by providing commonly used services such as access to citizen and property records, access control and authentication services, public key infrastructure, and support for digital signatures as reusable solution components and processes. By describing the components and data models at a higher semantic level, these are made easily accessible to an intuitive interface. Solution managers are thus able to use several aspects of the framework to customize solutions.

Tokyo Research Laboratory

The Tokyo Research Laboratory (TRL) has developed numerous core technologies for IBM products and solutions and has actively collaborated with other IBM research laboratories worldwide as well as with government research institutes, universities, and customers. TRL conducts research on a wide range of topics, including analytics and optimization, security and privacy, system design, server middleware, Internet technology, knowledge infrastructure, human–computer interactions and accessibility, optical technology, and future clients. For this journal issue, five subjects were selected to illustrate this wide scope of work.

Accessibility technology focuses on users who cannot access computers and information by means of standard interfaces. An overview of this work is provided in the paper by Maeda et al. Visually disabled people can utilize the Home Page Reader to access the Internet. It was developed at TRL and is now available in eight countries. This work is being extended to improve accessibility for blind, visually impaired, and elderly individuals. Recent work on a Web Accessibility Transcoding System for transforming complicated Web pages into accessible ones is also described.

On-demand design service innovations have been implemented through several customer engagements for LSI design. It has become possible to create a system with LSI that combines many system functions on a single chip by means of improvements in the technology for manufacturing semiconductors. Although complexity increases as circuit integration becomes large-scale, the time that can be spent on design development is becoming shorter, and so research on element technologies for the efficient design of systems is key. One such activity that uses advanced large-scale, highly efficient FPGAs (Field Programmable Gate Arrays) to verify ASIC designs in real time is described in the paper by Shimizu et al.

Java™ has become the mainstream language for Web applications. In its early days, because of its interpretive execution, it was notoriously slow. This issue has been resolved by Just-In-Time (JIT) compilers, which greatly

improve performance by dynamically translating byte code to native code. The development of JIT compilers and major components of JVMs™ (Java Virtual Machines) has been carried out since 1995; recent work is described by Suganuma et al. The JVMs and the JIT compilers have been shipped worldwide with all IBM platforms.

In the area of application development, the focus is on Web and pervasive application development tools. Page designers and Java programmers who have different roles and skills collaborate on the development of Web applications using a wide variety of technologies, such as HTML, JSP (JavaServer Pages™), and servlets. For large-scale applications it is important to create technologies and tools to minimize the task of maintaining the integrity of artifacts during development. A model-driven approach to the development of Web applications is described by Tai et al.

In the area of analytics and optimization, TRL has worked on many optimization problems, including supply chain management, logistics, air crew scheduling, and production line scheduling. Research on optimization algorithms strives to solve problems in large-scale production design and operations scheduling in manufacturing industries. The paper by Okano et al. describes how optimization technology was used successfully on a project for a steel company. Researchers in Japan and the United States collaborated with IBM Global Services personnel on large-scale optimization for operations involving production scheduling for steel coil products. TRL developed key algorithms for each process and for overall scheduling based on optimization expertise in solving traveling-salesman problems.

James Yeh

Director
IBM China Research Laboratory

Ponani S. Gopalakrishnan

Director
IBM India Research Laboratory

Kazushi Kuse

Director
IBM Tokyo Research Laboratory