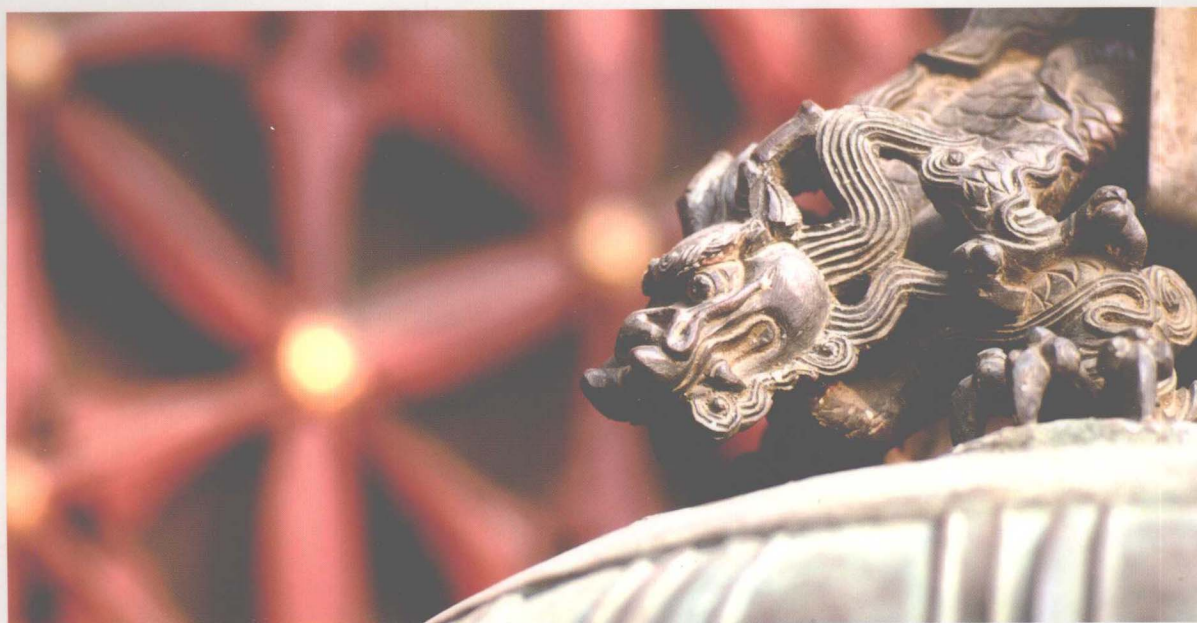
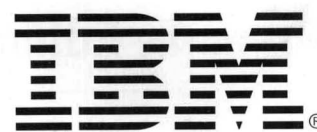




China Research Laboratory
Collection of
Selected Papers
2005-2008

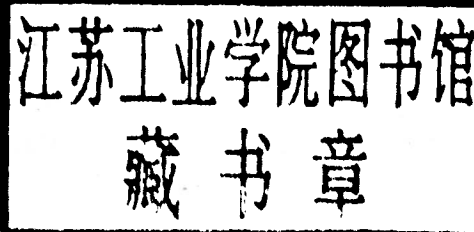


Tsinghua University Press



China Research Laboratory
Collection Of
Selected Papers
2005-2008

Editor-in-Chief:
Honesty Young



Tsinghua University Press
Beijing

China Research Laboratory Collection of Selected Papers 2005-2008/Honesty Young
© 2008 by Tsinghua University Press

All rights reserved. No part of this book may be reproduced,
In any form or by a means, without permission in writing from the publisher.

Tsinghua University Press
Beijing, 100084, China
www.tup.com.cn
ISBN: 978-7-302-18659-5
Price: \$80.00

版权所有，翻印必究。

图书在版编目 (CIP) 数据

IBM 中国研究院论文选集 (2005-2008): 英文/杨诚主编. —北京: 清华大学出版社, 2008.9
ISBN 978-7-302-18659-5

I. I … II. 杨… III. 计算机科学—文集—英文 IV. TP3-53

中国版本图书馆 CIP 数据核字 (2008) 第 146104 号

责任编辑: 冯昕
封面设计: 张敏
版式设计: 鲁静

出版发行: 清华大学出版社

<http://www.tup.com.cn>

社总机: 010-62770175

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

地址: 北京清华大学学研大厦 A 座

邮编: 100084

邮购: 010-62786544

印装者: 北京雅昌彩色印刷有限公司

开本: 210×285

版次: 2008 年 9 月第 1 版

定价: 80 美元

印张: 16

印次: 2008 年 9 月第 1 次印刷

From the Editors

IBM Research was originally established in 1945 in New York. During the past 63 years, IBM Research has remained a premier industrial research organization. As a globally integrated organization, IBM Research currently consists of eight laboratories throughout the world. These laboratories have a balanced portfolio of projects ranging from fundamental exploratory research to in-market research to address significant and challenging real world problems for our customers. Consequently, IBM Research has publications that cover both academic and industry specific topic areas. Since 1993, IBM has led every company in the number of patents granted in the United States. These patents are often followed by top quality papers.

Founded in 1995, the IBM China Research Laboratory (CRL) was the first research branch of a multinational corporation in China. It has attracted many top graduate students in China. In 2005, the 10th anniversary of CRL, a first volume of selected CRL-authored papers was published. Since then CRL has significantly enhanced its publication quality and quantity. This second volume contains 22 papers published in first-class conferences and journals. These selected papers represent about 6% of CRL papers since the first volume was published. During this period, CRL has further contributed to IBM's patent portfolio by significantly increasing the number of patent applications. Several CRL inventors have been named IBM master inventors. Often CRL collaborates with our world-wide IBM colleagues and with universities through the IBM university relations program. The result of this mutually beneficial collaboration is demonstrated by many papers having non-CRL co-authors. We were fortunate to have 15 world-class experts serve on our review board to help select the papers for this volume. They include two IBM Fellows, an IBM Distinguished Engineer, and a Fellow of the Chinese Academy of Sciences.

This volume has 11 parts which cover the major research fields at CRL. These fields range from infrastructure to solutions & services.

Analytics and Optimization

Analytics and Optimization is practiced throughout IBM, with the Research Division providing a focal point for activities in the science and the application of this field. The current application areas CRL is focused on include: facility location, network optimization, rule systems, yield management, supply chain modeling including environmental impacts, manufacturing planning and scheduling, transportation modeling. New initiatives include modeling and software environments for collaborative planning and water management.

Collaboration

Collaboration research at CRL is working on the creation of new and novel collaboration technologies, platforms, tools and middleware to improve business processes by enhancing personal and team productivity. Recent advancements in Information Management and Collaboration are the result of IBM research accomplishments. New collaboration technologies invented by CRL have been integrated and delivered in IBM products such as Websphere and OminFind. Current research is focused on domain specific social networking platforms, document-driven collaboration, and collaboration middleware.

Communications & Networking

The convergence of information technology (IT) and telecommunications is happening. This drives key technologies for the next generation network, server platform, and middleware to support converged data, voice, and web applications. Session Initiation Protocol (SIP) is a very good example of this convergence. SIP is a signaling protocol used for establishing and then disconnecting multimedia communication sessions such as voice and video calls over the Internet. It requires considering tradeoffs between system throughput and response time from traditional Internet protocols such as Hypertext Transfer Protocol (HTTP). Research at CRL has identified bottlenecks and devised several critical technologies, such as a SIP offloading engine and a front-end SIP flow management system. Both of these technologies significantly improve system performance. CRL developed prototypes and conducted extensive experiments which demonstrate that these novel techniques are generally applicable over a very broad range of SIP workloads. Other CRL research areas include converged network management, mobile data analytics, mobile communication enabled business processes, and telecom value added service creation & delivery.

Distributed Computing

CRL distributed computing research interests are on building cloud computing infrastructure with extremely low support costs for the delivery of web services by exploring new ideas in virtualization, green technology, automatic scaling, and management as a service. Other core technology interests are open and robust virtual world/3D Internet platforms with ultra-scalability, user created content management, virtual world interoperability, and integration with web.

From the Editors

Knowledge Management

Knowledge Management Research seeks to address challenging problems of advanced information and knowledge systems that are enterprise-wide to world-wide in scope. The mission is to raise the performance, scalability and usability to promote and demonstrate business value through real-world applications. Knowledge representation, data management, information retrieval, and semantic web are areas where CRL has a focused research interest. New results have been integrated into IBM products such as Information Management, Websphere, and Rational, and published and presented at conferences such as SIGMOD, VLDB, WWW, and ISWC.

Natural Language Processing

CRL research topics in Natural Language Processing include: named entity recognition, information extraction, summary generation, linguistic search, categorization, clustering, topic detection and tracking, duplicate and similar text detection, data cleansing, and semantic analysis. New research results have been published and presented at conferences such as SIGIR and WWW.

Service Science, Management and Engineering (SSME)

The global economy is changing. Service industries are becoming an important driving force for a sustainable economy. Presently service industries are transforming into a so called modern service industry through the use of IT. Although many countries have special programs and policies to promote the creation of a modern service industry, there is still a lack of large-scale well defined modern service practices. SSME research attempts to identify the methods, technologies, and policies needed to advance the development of a modern service industries. In order to realize this, CRL is working jointly with universities to collect, analyze, monitor, and verify models of the know-how for the development of service based industries.

Software & Service Engineering in Next Generation Services

Next Generation Services (NGS) focuses on business innovation and technology breakthroughs to enable the generation & operation of web delivered services, including Software as a Service (SaaS). The movement to a service economy is leading to the emergence of modern services which are being delivered through advanced web technologies. CRL NGS research is focused on technologies for web software engineering and web delivered service operations. These technologies create a low cost, efficient, and secure platform to enable the delivery and operation of web delivered services, as well

as the creation and composition of these services through collaborative web communities. Specific technical areas includes: SaaS delivery middleware, web security & trust technologies, web delivered service programming and composition, service operation excellence, software & service engineering, and service assets. The selected papers in this volume are a small representative sample of the work CRL has done.

Speech Technology

In the human-centric solution area, CRL aims to help user's interact with information & knowledge through intuitive user interfaces, and facilitate more pervasive interaction modalities. In the applied speech technology area, CRL has enabled technologies for conversational, multi-lingual and multi-modal interactions, including multimedia content management, advanced text to speech, automatic speech recognition/transcription, audio searching, melody searching, voice morphing, speech to speech translation, and speech analytics.

Systems

Computer Systems are the foundation on which end-to-end solutions and services are developed and deployed. Multicore architectures offer a new opportunity to expand Information Technology (IT) into new solution areas. However, efficiently making use of the available computing power provided by a multicore processor is a challenge for software engineers. CRL has developed technologies to hide the hardware complexity from software engineers without losing performance. A hybrid compile-time/runtime approach has been designed and implemented on an IBM Cell/B.E. multicore processor.

Increased processing power requires the network connecting the processors to increase bandwidth to maintain a balanced system. The embedded parallelism in a multicore processor provides a way to use IT for networking. A novel Internet Protocol (IP) parallel lookup scheme for a high speed network which can efficiently be implemented on a multicore processor has been developed. Previously, comparable functional capabilities required special hardware to achieve similar performance requirements.

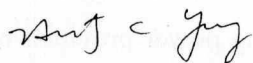
CRL research areas include highly scalable systems, networking, computer architecture and microarchitecture modeling, compilers, operating systems, stream computing, and end-to-end performance evaluation and optimization of applications. A few of these areas are presented in the papers selected for this volume.

From the Editors

Web Technology

Web Technology Research at CRL focuses on search and mining of unstructured, semi-structured and structured data on the web. As the web keeps quickly growing, a key challenge is information overloading. Effective mining of the web not only helps an individual user with knowledge acquisition, but is also a critical strategic necessity for many companies. Current CRL work on natural language processing is advancing the state of the art and on the frontier of web technologies. The traditional Web 1.0 involved the development of effective algorithms for page classification, data cleansing, sentiment mining, and competitor mining. With the fast-growing social Web 2.0, the new challenges are socially accumulated web data, such as social search, social browsing, and social community mining.

CRL has grown in many dimensions. One can appreciate this growth by the wide variety and depth of the papers selected for this volume. The publication of this second volume coincides with the opening of the IBM China Research Laboratory - Shanghai. Based on what CRL has already achieved, the future potential is unbounded. To learn more about CRL and obtain the latest information, please visit www.research.ibm.com/beijing/.



Honesty Young



J P Fasano



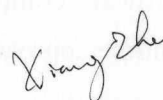
Yue Pan



Zhu Tang



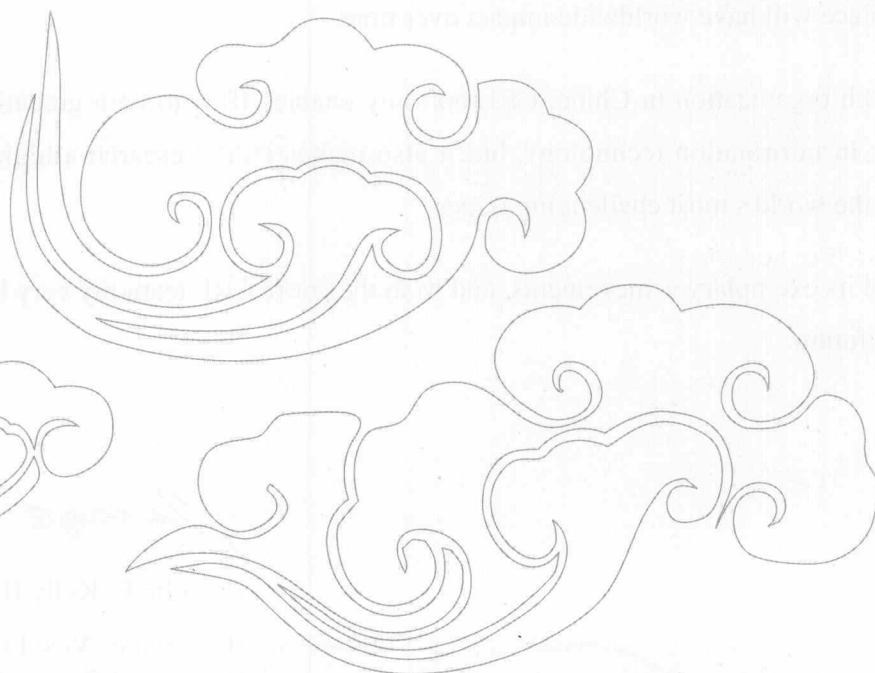
Qing Wang



Zhe Xiang

IBM China Research Laboratory

Preface



From Dr. John E. Kelly III

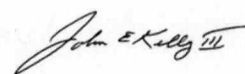
When it opened in 1995, IBM's China Research Laboratory (CRL) was the first-ever research organization to be established by a multinational company in China. And this year, with the opening of a satellite laboratory in Shanghai, CRL has significantly expanded its coverage in China. Over the years, CRL has attained many notable achievements, and many of them are reflected in this very special collection of publications. Though the papers published in this book represent only a fraction of CRL's broad project portfolio, each reflects the richness and diversity of CRL's disciplines, approaches, partnerships, client relationships and results.

The papers cover a wide range of topics -- from bold exploratory research to marketable technology and solutions. As you read them, you also will note that non-CRL co-authors confirm the CRL team's ability to leverage the benefits of collaboration with universities, governments and other partners, both inside and outside of IBM, which is so important in driving innovation in the 21st century. Overall, the breadth and depth of these papers demonstrate the outstanding caliber of the CRL team.

Today, China's economy is among the largest in the world, representing a significant portion of the world's GDP growth. Recognizing China's importance to the global economy, IBM formed an Emerging Markets organization headquartered in Shanghai. Through this organization, IBM researchers experiment in the marketplace to more fully understand China's unique societal and industrial issues and create powerful new solutions. We fully expect that most of the CRL technologies and solutions inspired by the China marketplace will have worldwide impact over time.


As the leading research organization in China, CRL not only enables IBM to be a globally-integrated enterprise and leader in information technology, but it also makes IBM Research a highly-desirable partner in addressing the world's most challenging issues.

I congratulate CRL on its exemplary achievements, and wish the entire CRL team my very best for even greater success in the future.



Dr. John E. Kelly III

IBM Senior Vice President and
Director of Research



From Dr. James Yeh

As I witnessed IBM China Research Laboratory grew from a small laboratory of about fifty people ten years ago to a premier lab now of over two hundred researchers, it is gratifying to see the set of outstanding publications that the laboratory produced from its excellent research. The research topics that China Research Laboratory currently is conducting cover a wide range of areas that are forefronts of Computer Science and Service Science. The research serves the role of defining the world of the future in one hand and explore how the technologies can bring further value to enterprises and individuals in the other.

This Book of Publications only reflects a small portion of the all the lab publications. Nevertheless it gives you a glimpse of the future and the depth of the research that the lab is doing. Let me congratulate the members of China Research Laboratory on your leadership in demonstrating the professionalism and in creating the knowledge that is crucial in moving the world forward.



Dr. James Yeh
Chief Technology Officer
IBM Greater China Group

From Dr. Thomas Li

When we published the first IBM China Research Laboratory (CRL) Paper Collection in 2005, it was for the 10th anniversary celebration of CRL. Today, we publish the second Paper Collection to mark a major milestone, the establishment of IBM China Research Laboratory – Shanghai. It is a strong evidence of IBM commitment to root, nurture, and flourish the advanced research in information technology in China.

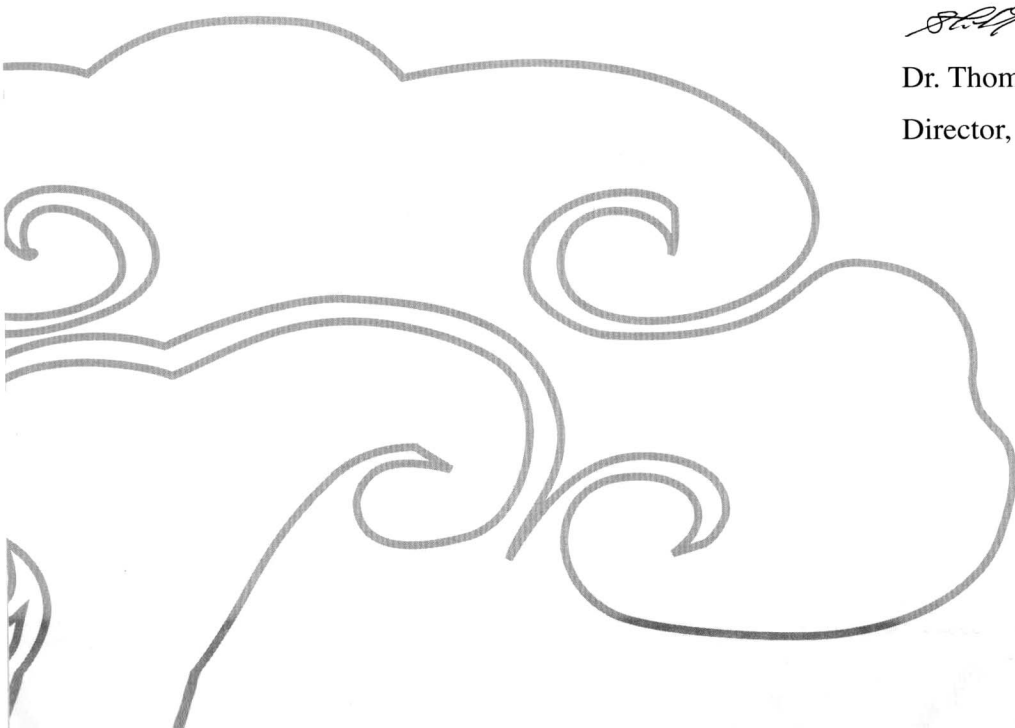
As one of the eight IBM Research laboratories worldwide and the pioneer of research institutes in China, CRL has successfully developed new technologies through collaborations with global technical communities in many fronts. We are putting together a collection of our papers from key journals and conference proceedings from 2005 to 2008. Though this collection represents only a small percentage of the papers we published, it reflects the breadth and depth of our research and innovation during the past three years. The papers together also illustrate the foundation to support our further research advancement in transforming web, which can elevate user experience with emerging services via an integrated infrastructure of conventional internet, mobile, and wireless.

In this paper collection, you may find the traces of cross-pollination among multiple disciplines, including Computer Science, Human-Centric Computing, and Management Science & Operations Research. They show the key characteristics of CRL for over 13 years: Open, Collaboration, and Innovation, which are the basic elements to foster a good environment for world-class research that can contribute to a better world.



Dr. Thomas Li

Director, IBM China Research Laboratory



为IBM中国研究院研究文集而作

IBM作为一家著名跨国公司,非常重视在中国的发展。自1995年在北京成立以来,IBM中国研究院已迅速发展成为IBM全球最重要的八大研究中心之一,在中文信息处理、数字媒体、普适计算和网格计算、知识管理、物流、商务管理与优化等方面取得了丰硕的研究成果。

近年来,中国经济发展面临着重要的结构调整,大力发展现代服务业。早在2005年,IBM就积极倡导服务科学的学科建设和科学研究,与国家教育部展开全面合作,和清华大学、北京大学等一批中国高校一起,成立联合研究机构,开设新的课程,选择共同感兴趣的课题进行研究,致力于推动现代服务业在中国的发展。在这些工作中,IBM中国研究院扮演了非常重要的角色。可以想见,IBM中国研究院在服务科学方面的努力,将为中国经济的发展做出巨大贡献。

随着IBM中国研究院的发展,他们在上海成立了分中心。这也标志着IBM中国研究院的进一步壮大。为了庆贺分中心的成立,他们编辑出版这一文集,收集了近3年来IBM研究中心在分析、优化、协同、通讯、网络、分布计算、知识管理、自然语言处理、服务、软件与方案工程、语音与用户交互技术、系统和互联网技术等个研究方面的成果。与3年前出版的文集相比,这些成果无论在问题的新颖性还是研究深度上都有较大提高,有着鲜明的理论联系实际的特点,非常值得学术同行们学习借鉴。我相信,这一文集的出版,必将对国内相关领域的研究起到很好的促进作用。

祝愿IBM中国研究院蓬勃发展,祝愿他们与中国高校的合作不断深入!



李衍达

中国科学院院士,清华大学教授





Contents

	Part I. Analytics & Optimization
03	Classification by Discriminative Regularization Bin Zhang, Fei Wang, Ta-Hsin Li, Wenjun Yin, Jin Dong
09	Software as Service Pricing: A Game Theory Perspective Chunhua Tian, Yan Zheng, Zhongbo Jiang, Rongzeng Cao, Wei Sun, Wei Ding
	Part II. Collaboration
15	Group Efficacy in Asynchronous vs. Multi-synchronous Virtual Teams: An Empirical Study Yingxin Pan, Chen Zhao
29	Hidden Sentiment Association in Chinese Web Opinion Mining Qi Su, Xinying Xu, Honglei Guo, Zhili Guo, Xian Wu, Xiaoxun Zhang, Bin Swen, Zhong Su
	Part III. Communications & Networking
41	Flow Management for SIP Application Servers Jing Sun, Jinfeng Hu, Ruixiong Tian, Bo Yang
49	SIP Parsing Offload: Design and Performance Jia Zou, Wei Xue, Zhiyong Liang, Yixin Zhao, Bo Yang, Ling Shao
	Part IV. Distributed Computing
57	Modeling and Verifying Configuration in Service Deployment Ying Li, Jie Qiu, Kewei Sun, Ying Chen
65	Research and Implementation of Knowledge-enhanced Information Services Bo Yang, Hao Wang, Liang Liu, Qian Ma, Ying Chen, Hui Lei
	Part V. Knowledge Management
79	Explorations in the Use of Semantic Web Technologies for Product Information Management Jean-Sébastien Brunner, Li Ma, Chen Wang, Lei Zhang, Daniel C. Wolfson, Yue Pan, Kanvitha Srinivas
89	Semplore: An IR Approach to Scalable Hybrid Query of Semantic Web Data Lei Zhang, Qiaoling Liu, Jie Zhang, Haofen Wang, Yue Pan, Yong Yu
	Part VI. Natural Language Processing
105	An Integrated System for Building Enterprise Taxonomies Li Zhang, Tao Li, Shixia Liu, Yue Pan



- 131 **Floatcascade Learning for Fast Imbalanced Web Mining**
Xiaoxun Zhang, Xueying Wang, Honglei Guo, Zhili Guo, Xian Wu, Zhong Su
- Part VII. Services**
- 143 **A Dynamic Concept Interpretation Approach for OWL Extension**
Shun Jiang, Chunhua Tian, Feng Li, Hao Zhang, Wei Ding
- 149 **Services Pricing through Business Value Modeling and Analysis**
Wei Ding
- Part VIII. Software & Service Engineering in Next Generation Services**
- 159 **BPEL-Unit: JUnit for BPEL Processes**
Zhongjie Li, Wei Sun
- 171 **Towards Facilitating Development of SOA Application with Design Metrics**
Wei Zhao, Ying Liu, Jun Zhu, Hui Su
- Part IX. Speech & User Interface Technologies**
- 185 **Voice Conversion by Combining Frequency Warping with Unit Selection**
Zhiwei Shuang, Fanping Meng, Yong Qin
- 189 **Voice-Melody Transcription under a Speech Recognition Framework**
Danning Jiang, Michael Picheny, Yong Qin
- Part X. Systems**
- 195 **ATCAM-Based Distributed Parallel IP Lookup Scheme and Performance Analysis**
Kai Zheng, Chengchen Hu, Hongbin Lu, Bin Liu
- 209 **Orchestrating Data Transfer for the Cell/B.E. Processor**
Tong Chen, Haibo Lin, Tao Zhang, Kathryn M. O'Brien, John Kevin O'Brien
- Part XI. Web Technology**
- 221 **Exploring Folksonomy for Personalized Search**
Shengliang Xu, Shenghua Bao, Ben Fei, Zhong Su, Yong Yu
- 229 **Towards Effective Browsing of Large Scale Social Annotations**
Rui Li, Shenghua Bao, Ben Fei, Zhong Su, Yong Yu

PART I.

Analytics & Optimization

