

# COMMUNICATIONS

CACM.ACM.ORG

OF THE

# ACM

12/08 VOL.51 NO.12

## Surviving the Data Deluge

Open Information  
Extraction  
from the Web

CTOs on  
Virtualization

Living Machines

High-Performance  
Web Sites



## Departments

- 5 **Publications Board Letter**  
**ACM Publications: Access and Sustainability**  
*By Ronald F. Boisvert and Holly Rushmeier*
- 
- 7 **Publisher's Corner**  
**Looking Back and Forward**  
*By Scott E. Delman*
- 
- 8 **Letters To The Editor**  
**Apple Builds Great Platforms, Too, Not Just Products**
- 
- 10 **CACM Online**  
**Previewing the New CACM Web Site**  
*By David Roman*
- 
- 31 **Calendar**
- 
- 96 **Careers**

## Last Byte

- 118 **Puzzled**  
**Solutions and Sources**  
*By Peter Winkler*
- 
- 120 **Future Tense**  
**Pester Power**  
*By Cory Doctorow*

## News

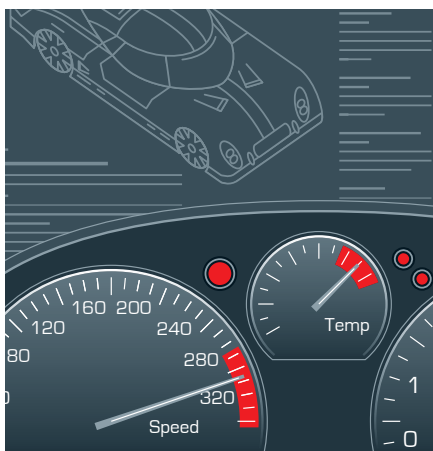


- 11 **Living Machines**  
 Researchers of molecular computing and communication are focusing on the type of breakthroughs that are needed to make the vision of ultrasmall, biocompatible computers a reality.  
*By Kirk L. Kroeker*
- 
- 14 **Touching the Future**  
 In combination with finger and hand gestures, multitouch input is enabling users to manipulate and display information in innovative and unprecedented ways.  
*By Ted Selker*
- 
- 17 **Upwardly Mobile**  
 Mobile phones are bridging the digital divide and transforming many economic, social, and medical realities, particularly in developing nations.  
*By Samuel Greengard*
- 
- 20 **Making a Difference**  
 The Grace Hopper Celebration featured technical talks, workshops, networking events, and lively discussions about increasing the number of women in computer science.  
*By Leah Hoffmann*

## Viewpoints

- 21 **Emerging Markets**  
**Labor Supply in the Indian IT Industry**  
 Exploring the evolving dynamics and interconnectedness of India's educational system and its IT work force.  
*By Alok Aggarwal*
- 
- 24 **IT Policy**  
**Advising Policymakers Is More Than Just Providing Advice**  
 What are the factors that make certain advisory committee reports successful while others are not?  
*By Cameron Wilson and Peter Harsha*
- 
- 27 **Kode Vicious**  
**Permanence and Change**  
 Highlighting the importance of doing one's best in view of code longevity and the impermanence of the changineer.  
*By George V. Neville-Neil*
- 
- 29 **The Profession of IT**  
**Evolutionary System Development**  
 Large systems projects are failing at an alarming rate. It's time to take evolutionary design methods off the shelf.  
*By Peter J. Denning, Chris Gunderson, and Rick Hayes-Roth*
- 
- 32 **Interview**  
**Database Dialogue with Pat Selinger**  
 Relational database pioneer Patricia G. Selinger explores the vast realm of database technology and trends in a wide-ranging discussion with Microsoft's James Hamilton.  
*By James Hamilton*

## Practice



- 36 **High-Performance Web Sites**  
Want to make your Web site fly?  
Focus on frontend performance.  
*By Steve Souders*

- 43 **CTO Virtualization Roundtable, Part II**  
When it comes to virtualization  
platforms, experts say focus first  
on the services to be delivered.  
*By Mache Creeger, Moderator*



**About the Cover:** To create an image of a wave of data, German artist Thomas Herbrich photographed over 1,000 printed pages then reduced them to 6cm through multiple photocopying. He then began the arduous task of adhering thousands of these miniature pages onto corrugated Styrofoam sheets to create and photograph this 3D scene.

The effect of flying papers at the top of the wave was created by throwing hundreds of mini-pages out of a bucket and photographing them. He then blended the separate images together to form the wave. For more about his work, see <http://www.herbrich.com/>.

## Contributed Articles

- 50 **Got Data? A Guide to Data Preservation in the Information Age**  
Tools for surviving a data deluge to ensure your data will be there when you need it.  
*By Francine Berman*
- 58 **Ontologies and the Semantic Web**  
How ontologies provide the semantics behind the Semantic Web, as explained here with the help of Harry Potter and his owl Hedwig.  
*By Ian Horrocks*

## Review Articles

- 68 **Open Information Extraction from the Web**  
Targeted IE methods are transforming into open-ended techniques.  
*By Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S. Weld*

## Research Highlights

- 75 **Technical Perspective**  
**One Size Fits All: An Idea Whose Time has Come and Gone**  
*By Michael Stonebraker*
- 77 **Breaking the Memory Wall in MonetDB**  
*By Peter A. Boncz, Martin L. Kersten, and Stefan Manegold*
- 86 **Technical Perspective**  
**Patching Program Errors**  
*By Martin C. Rinard*
- 87 **Exterminator: Automatically Correcting Memory Errors with High Probability**  
*By Gene Novark, Emery D. Berger, and Benjamin G. Zorn*

## Virtual Extension

As with all magazines, page limitations often prevent the publication of articles that might otherwise be included in the print edition. To ensure timely publication, ACM created *Communications'* Virtual Extension (VE).

VE articles undergo the same rigorous review process as those in the print edition and are accepted for publication on their merit. These articles are now available to ACM members in the Digital Library.

**Enhancing Business Performance via Vendor-Managed Inventory Applications**  
*Peter Duchessi and InduShobha Chengalur-Smith*

**A Framework of ICT Exploitation for E-Participation Initiatives**  
*Chee Wei Phang and Atreyi Kankanhalli*

**Two Approaches to an Information Security Library**  
*Michael Näf and David Basin*

**Opening Up to Agile Games Development**  
*Patrick Stacey and Joe Nandhakumar*

**Worst Practices in Search Engine Optimization**  
*Ross A. Malaga*

**From Genesis to Revelations: The Technology Diaspora**  
*Pierre Berthon, Leyland F. Pitt, and Richard T. Watson*

**Evolving Innovations through Design and Use**  
*Michael J. Davern and Carla L. Wilkin*

**Technical Opinion**  
**Bosses and their E-Manners**  
*Vivien K.G. Lim, Thompson S.H. Teo, and Jen Yui Chin*



Association for Computing Machinery  
Advancing Computing as a Science & Profession



ACM, the world's largest educational and scientific computing society, delivers resources that advance computing as a science and profession. ACM provides the computing field's premier Digital Library and serves its members and the computing profession with leading-edge publications, conferences, and career resources.

#### Executive Director and CEO

John White

#### Deputy Executive Director and COO

Patricia Ryan

#### Director, Office of Information Systems

Wayne Graves

#### Director, Office of Financial Services

Russell Harris

#### Director, Office of Membership

Lillian Israel

#### Director, Office of Publications

Mark Mandelbaum

#### Director, Office of SIG Services

Donna Cappel

#### ACM COUNCIL

##### President

Wendy Hall

##### Vice-President

Alain Chenais

##### Secretary/Treasurer

Barbara Ryder

##### Past President

Stuart I. Feldman

##### Chair, SGB Board

Alexander Wolf

##### Co-Chairs, Publications Board

Ronald Boisvert, Holly Rushmeier

##### Members-at-Large

Carlo Ghezzi;

Anthony Joseph;

Mathai Joseph;

Kelly Lyons;

Bruce Maggs;

Mary Lou Soffa;

##### SGB Council Representatives

Norman Jouppi;

Robert A. Walker;

Jack Davidson

#### PUBLICATIONS BOARD

##### Co-Chairs

Ronald F. Boisvert and Holly Rushmeier

##### Board Members

Gul Agha; Michel Beaudouin-Lafon;

Jack Davidson; Carol Hutchins;

Ee-ping Lim; M. Tamer Ozsu; Vincent Shen;

Mary Lou Soffa; Ricardo Baeza-Yates

#### ACM U.S. Public Policy Office

Cameron Wilson, Director

1100 Seventeenth St., NW, Suite 507

Washington, DC 20036 USA

T (202) 659-9711; F (202) 667-1066

#### Computer Science Teachers Association

Chris Stephenson

Executive Director

2 Penn Plaza, Suite 701

New York, NY 10121-0701 USA

T (800) 401-1799; F (541) 687-1840

#### Association for Computing Machinery (ACM)

2 Penn Plaza, Suite 701

New York, NY 10121-0701 USA

T (212) 869-7440; F (212) 869-0481

# COMMUNICATIONS OF THE ACM

A monthly publication of ACM Media

*Communications of the ACM* is the leading monthly print and online magazine for the computing and information technology fields. *Communications* is recognized as the most trusted and knowledgeable source of industry information for today's computing professional. *Communications* brings its readership in-depth coverage of emerging areas of computer science, new trends in information technology, and practical applications. Industry leaders use *Communications* as a platform to present and debate various technology implications, public policies, engineering challenges, and market trends. The prestige and unmatched reputation that *Communications of the ACM* enjoys today is built upon a 50-year commitment to high-quality editorial content and a steadfast dedication to advancing the arts, sciences, and applications of information technology.

## STAFF

### GROUP PUBLISHER

Scott E. Delman

[publisher@cacm.acm.org](mailto:publisher@cacm.acm.org)

### Executive Editor

Diane Crawford

### Managing Editor

Thomas E. Lambert

### Senior Editor

Andrew Rosenbloom

### Senior Editor/News

Jack Rosenberger

### Web Editor

David Roman

### Editorial Assistant

Zarina Strakhan

### Rights and Permissions

Deborah Cotton

### Art Director

Andrij Borys

### Associate Art Director

Alicia Kubista

### Assistant Art Director

Mia Angelica Balaquiot

### Production Manager

Lynn D'Addesio

### Director of Media Sales

Jonathan Just

### Advertising Coordinator

Jennifer Ruzicka

### Marketing & Communications Manager

Brian Hebert

### Public Relations Coordinator

Virginia Gold

### Publications Assistant

Emily Eng

### Columnists

Alok Aggarwal; Phillip G. Armour;

Martin Campbell-Kelly;

Michael Cusumano; Peter J. Denning;

Shane Greenstein; Mark Guzdial;

Peter Harsha; Leah Hoffmann;

Mari Sako; Pamela Samuelson;

Gene Spafford; Cameron Wilson

## CONTACT POINTS

### Copyright permission

[permissions@cacm.acm.org](mailto:permissions@cacm.acm.org)

### Calendar items

[calendar@cacm.acm.org](mailto:calendar@cacm.acm.org)

### Change of address

[acmcoa@cacm.acm.org](mailto:acmcoa@cacm.acm.org)

### Letters to the Editor

[letters@cacm.acm.org](mailto:letters@cacm.acm.org)

## WEB SITE

<http://cacm.acm.org>

## AUTHOR GUIDELINES

<http://cacm.acm.org/guidelines>

## ADVERTISING

### ACM ADVERTISING DEPARTMENT

2 Penn Plaza, Suite 701, New York, NY

10121-0701

T (212) 869-7440

F (212) 869-0481

### Director of Media Sales

Jonathan M. Just

[jonathan.just@acm.org](mailto:jonathan.just@acm.org)

**Media Kit** [acmm mediasales@acm.org](mailto:acmm mediasales@acm.org)

## EDITORIAL BOARD

### EDITOR-IN-CHIEF

Moshe Y. Vardi

[vic@cacm.acm.org](mailto:vic@cacm.acm.org)

## NEWS

### Co-chairs

Marc Najork and Prabhakar Raghavan

### Board Members

Brian Bershad; Hsiao-Wuen Hon;

Mei Kobayashi; Rajeev Rastogi;

Jeannette Wing

## VIEWPOINTS

### Co-chairs

William Aspray;

Susanne E. Hambrusch;

J Strother Moore

### Board Members

Stefan Bechtold; Judith Bishop;

Peter van den Besselaar; Soumitra Dutta;

Peter Freeman; Seymour Goodman;

Shane Greenstein; Mark Guzdial;

Richard Heeks; Susan Landau;

Carlos Jose Pereira de Lucena;

Helen Nissenbaum; Beng Chin Ooi

## PRACTICE

### Chair

Stephen Bourne

### Board Members

Eric Allman; Charles Beeler;

David J. Brown; Bryan Cantrill;

Terry Coatta; Mark Compton;

Benjamin Fried; Pat Hanrahan;

Marshall Kirk McKusick;

George Neville-Niel

The Practice section of the CACM

Editorial Board also serves as

the Editorial Board of ACM Queue.

## CONTRIBUTED ARTICLES

### Co-chairs

Al Aho and George Gottlob

### Board Members

Yannis Bakos; Gilles Brassard; Peter

Buneman; Andrew Chien; Anja Feldmann;

Blake Ives; Takeo Kanade; James Larus;

Igor Markov; Gail C. Murphy; Shree Nayar;

Lionel M. Ni; Sriram Rajamani; Avi Rubin;

Abigail Sellen; Ron Shamir; Larry Snyder;

Wolfgang Wahlster; Andy Chi-Chih Yao;

Willy Zwaenepoel

## RESEARCH HIGHLIGHTS

### Co-chairs

David A. Patterson and

Stuart J. Russell

### Board Members

Martin Abadi; P. Anandan; Stuart K. Card;

Deborah Estrin; Stuart I. Feldman;

Shafi Goldwasser; Maurice Herlihy;

Norm Jouppi; Andrew B. Kahng; Linda

Petzold; Michael Reiter;

Mendel Rosenblum; Ronitt Rubinfeld;

David Salesin; Lawrence K. Saul;

Guy Steele, Jr.; Gerhard Weikum;

Alexander L. Wolf

## WEB

### Co-chairs

Marti Hearst and James Landay

### Board Members

Jason I. Hong; Jeff Johnson;

Greg Linden; Wendy E. MacKay;

Jian Wang



BPA Audit Pending

## ACM Copyright Notice

Copyright © 2008 by Association for Computing Machinery, Inc. (ACM). Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or fee. Request permission to publish from [permissions@acm.org](mailto:permissions@acm.org) or fax (212) 869-0481.

For other copying of articles that carry a code at the bottom of the first or last page or screen display, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center; [www.copyright.com](http://www.copyright.com).

## Subscriptions

Annual subscription cost is included in the society member dues of \$99.00 (for students, cost is included in \$42.00 dues); the nonmember annual subscription rate is \$100.00.

## ACM Media Advertising Policy

*Communications of the ACM* and other ACM Media publications accept advertising in both print and electronic formats. All advertising in ACM Media publications is at the discretion of ACM and is intended to provide financial support for the various activities and services for ACM members. Current Advertising Rates can be found by visiting <http://cacm.acm.org/advertising> or by contacting ACM Media Sales at (212) 626-0654.

## Single Copies

Single copies of *Communications of the ACM* are available for purchase. Please contact [acmhlp@acm.org](mailto:acmhlp@acm.org).

## COMMUNICATIONS OF THE ACM

(ISSN 0001-0782) is published monthly by ACM Media, 2 Penn Plaza, Suite 701, New York, NY 10121-0701. Periodicals postage paid at New York, NY 10001, and other mailing offices.

## POSTMASTER

Please send address changes to *Communications of the ACM* 2 Penn Plaza, Suite 701 New York, NY 10121-0701 USA



Association for Computing Machinery

Printed in the U.S.A.



# research highlights

---

P. 76

**Technical  
Perspective**  
**One Size Fits All:  
An Idea Whose Time  
has Come and Gone**

By Michael Stonebraker

P. 77

**Breaking the Memory  
Wall in MonetDB**

By Peter A. Boncz, Martin L. Kersten, and Stefan Manegold

---

P. 86

**Technical  
Perspective**  
**Patching  
Program Errors**

By Martin C. Rinard

P. 87

**Exterminator: Automatically  
Correcting Memory Errors  
with High Probability**

By Gene Novark, Emery D. Berger, and Benjamin G. Zorn