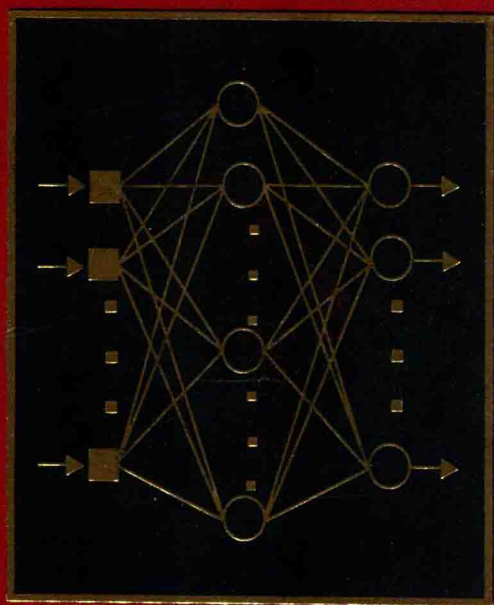


Advances in

COMPUTERS

Volume **64**

New Programming Paradigms



Edited by

MARVIN V. ZELKOWITZ

Advances in **COMPUTERS**

New Programming Paradigms

EDITED BY

MARVIN V. ZELKOWITZ

Department of Computer Science
and Institute for Advanced Computer Studies
University of Maryland
College Park, Maryland

VOLUME 64



ELSEVIER
ACADEMIC
PRESS

Amsterdam Boston Heidelberg London New York Oxford
Paris San Diego San Francisco Singapore Sydney Tokyo

ELSEVIER B.V.
Radarweg 29
P.O. Box 211, 1000 AE Amsterdam
The Netherlands

ELSEVIER Inc.
525 B Street, Suite 1900
San Diego, CA 92101-4495
USA

ELSEVIER Ltd
The Boulevard, Langford Lane
Kidlington, Oxford OX5 1GB, UK

ELSEVIER Ltd
84 Theobalds Road
London WC1X 8RR, UK

© 2005 Elsevier Inc. All rights reserved.

This work is protected under copyright by Elsevier Inc., and the following terms and conditions apply to its use:

Photocopying

Single photocopies of single chapters may be made for personal use as allowed by national copyright laws. Permission of the Publisher and payment of a fee is required for all other photocopying, including multiple or systematic copying, copying for advertising or promotional purposes, resale, and all forms of document delivery. Special rates are available for educational institutions that wish to make photocopies for non-profit educational classroom use.

Permissions may be sought directly from Elsevier's Rights Department in Oxford, UK: phone (+44) 1865 843830, fax (+44) 1865 853333, e-mail: permissions@elsevier.com. Requests may also be completed on-line via the Elsevier homepage (<http://www.elsevier.com/locate/permissions>).

In the USA, users may clear permissions and make payments through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA; phone: (+1) (978) 7508400, fax: (+1) (978) 7504744, and in the UK through the Copyright Licensing Agency Rapid Clearance Service (CLARCS), 90 Tottenham Court Road, London W1P 0LP, UK; phone: (+44) 20 7631 5555, fax: (+44) 20 7631 5500. Other countries may have a local reprographic rights agency for payments.

Derivative Works

Tables of contents may be reproduced for internal circulation, but permission of the Publisher is required for external resale or distribution of such material. Permission of the Publisher is required for all other derivative works, including compilations and translations.

Electronic Storage or Usage

Permission of the Publisher is required to store or use electronically any material contained in this work, including any chapter or part of a chapter.

Except as outlined above, no part of this work may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the Publisher.

Address permissions requests to: Elsevier's Rights Department, at the fax and e-mail addresses noted above.

Notice

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.

First edition 2005

ISBN: 0-12-012164-6

ISSN (Series): 0065-2458

© The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

Printed in the USA.

Advances

**in COMPUTERS
VOLUME 64**

Contributors

Dr. Coskun Bayrak is an associate professor in the Department of Computer Science at the University of Arkansas at Little Rock. He also has been serving on the Faculty of Applied Science since 2000. His primary research is in the intersection of software engineering, component based development, distributed systems, and data-mining. However, he also has interest in modeling and simulation, cellular automata, and monitoring and control. Dr. Bayrak has published over forty research articles in scientific conferences and journals, given tutorials at major conferences, and served on program committees for numerous international conferences and symposiums. Dr. Bayrak holds a BS from Slippery Rock University, and a MS from Texas Tech University, and Ph.D. from Southern Methodist University in Computer Science. He can be reached at cxbayrak@ualr.edu.

Professor Hal Berghel is Director of the School of Computer Science at the University of Nevada at Las Vegas, and Director of both the Center for Cybermedia Research and Identity Theft and Financial Fraud Research and Operations Center. He has held a variety of research and administrative positions in industry and academia during his twenty-five year career in computing. His current research focuses on Internet security and forensics, interactive and participatory computing environments, and dynamic models for cyberpublishing. His research work appears frequently in a variety of scientific and technical venues, and his columns, editorials, and articles appear regularly in such publications as Computer and the Communications of the ACM. He is a Fellow of both the ACM and IEEE. He can be reached at hlb@acm.org.

Dr. Abdur Chowdhury is the Chief Architect for the Search and Navigation Group at AOL. Additionally, he is an affiliated researcher with the Illinois Institute of Technology IR lab. Dr. Chowdhury has over 10 years of research and development experience in computer science. He has worked as both a researcher and a developer, which has given him a great perspective on balancing software design with innovation. During this time he has seven patent applications filed and over 55 publications in magazines, scientific conferences, journals and book chapters covering

many computer science topics like networking, operating systems, system scaling, and information retrieval. He can be reached at cabdur@aol.com.

Mr. Chad Davis is a software consultant. Typically, this work amounts to building proprietary software, for large corporations, with open source development tools. He is also a writer who only begrudgingly expends his literary energy on computer related topics, preferring by nature the much less practical forms of fiction and poetry. His research interests include distributed systems, object oriented design, and the open source culture. He holds graduate degrees in both Computer Science and Literature. He can be reached at chadmichael@excite.com.

Peter Gregor is a Senior Lecturer in Applied Computing at the University of Dundee in Scotland. His background is in psychology, community education and computing. He has a particular interest in making technology work for people and as such has a special interest in all practical aspects of human computer interaction and the design of effective and enjoyable user experience. He has published widely and pursues research actively, with funding from government, charities and commercial sources. He founded the Digital Media Access Group, a research and consultancy group focusing on inclusive design of technology, in 1999 and is its Director. He can be reached at pgregor@computing.dundee.ac.uk.

Professor John Harrison is an Assistant Professor in the School of Computer Science at the University of Nevada, Las Vegas. He conducts research within the Center for Cybermedia Research, where he serves as Associate Director of the Internet Forensic Laboratory and as a member of the Research Faculty of the Identity Theft and Financial Fraud Research Center. Prior to this he was a systems architect at Adspace Networks (Burlingame, CA), an information management consultant to the casino gaming industry and a tenured faculty member in the School of Information Technology and Electrical Engineering at the University of Queensland, Brisbane, Australia. He can be reached at jh@cs.unlv.edu.

Professor P.K. Kannan is Harvey Sanders Associate Professor of Marketing at the Robert H. Smith School of Business at the University of Maryland. He is the Associate Director for the Center for Excellence in Service. His current research stream focuses on new product development, pricing of information products and product lines, customer relationship management, and online and offline loyalty, and spatial modeling. His research papers have been published in *Management Science*, *Marketing Science*, *Journal of Academy of Marketing Science*, and *Communications of the ACM* among others. Dr. Kannan is a Fellow of the American Marketing Association Consortium on e-Commerce (2001). He is an Associate Editor for *Decision Support*

Systems and serves on the editorial boards of the *Journal of Service Research*, *Journal of Business Research*, and the *International Journal of Electronic Commerce*. Dr. Kannan has served as the Chair for the American Marketing Association Special Interest Group on Marketing Research. He is the recipient of the Krowe Award for Teaching Excellence (2001). He can be reached at pkannan@rhsmith.umd.edu.

Professor Alan Newell has been researching into computer systems to assist people with disabilities for over thirty years, and has published widely in this field. He was awarded the Lloyd of Kilgerran Prize from the Foundation for Science and Technology for this research, and has been appointed a Member of the Order of the British Empire for services to IT and communication for people with disabilities. He is also a Fellow of the Royal Society of Edinburgh, the British Computer Society, the Institution of Electrical Engineers, and an Honorary Fellow of the Royal College of Speech and Language Therapists.

Anupama D. Ramachandran is a doctoral student at the University of Maryland. He can be reached at aramacha@rhsmith.umd.edu.

Professor Roland T. Rust holds the David Bruce Smith Chair in Marketing at the Robert H. Smith School of Business at the University of Maryland, where he is the Chair of the Marketing Department and directs the Center for e-Service. His lifetime achievement honors include the American Marketing Association's Gilbert A. Churchill Award for contributions to marketing research, the Outstanding Contributions to Research in Advertising award from the American Academy of Advertising, the AMA's Career Contributions to the Services Discipline Award, Fellow of the American Statistical Association, and the Henry Latané Distinguished Doctoral Alumnus Award from the University of North Carolina at Chapel Hill. He has won best article awards for articles in *Marketing Science*, *Journal of Marketing Research*, *Journal of Marketing*, *Journal of Advertising*, and *Journal of Retailing*, as well as the Marketing science Institute (MSI) Best Paper Award twice. His seven books include *e-Service*, *Driving Customer Equity* (winner of the 2002 Berry-AMA Book Prize for the best marketing book of the previous three years), *Service Marketing* and *Return on Quality*. He is the founder and Chair of the AMA Frontiers in Services Conference, and serves as founding Editor of the *Journal of Service Research*. He can be reached at rrust@rhsmith.umd.edu.

Professor Debashis Saha is currently a Full Professor in the MIS and Computer Science Group of Indian Institute of Management—Calcutta (IIM-C) in India. Previously, he was with Jadavpur University in the Computer Science & Engg. Department from 1990 to 2001. His major research interests include Pervasive Commu-

nication and Computing, Wireless Networking and Mobile Computing, and WDM optical networking. He has co-authored five books and a monograph. He has published more than 150 papers in various conferences and journals. He has received his Bachelor degree from Jadavpur University, Calcutta, and his Masters and Ph.D. degrees from IIT at Kharagpur in India. Dr. Saha is the founder chair of Calcutta Chapter of IEEE Communications Society, a Senior Life Member of Computer Society of India, Senior Member of IEEE, and a member of IFIP WG 6.8 and WG 6.10. He can be reached at ds@iimcal.ac.in.

Sang Shin is a Java Technology Evangelist and Architect at Sun Microsystems, Inc. He has developed and delivered numerous talks on J2EE, EJB, JMS, J2ME, Web application frameworks and Web services technologies to worldwide developer audiences. He also teach software engineering courses in Brandeis University. He can be reached at sang.shin@sun.com.

David Sloan is a founder member and project lead of the Digital Media Access Group, a research and consultancy group focusing on inclusive design of technology, based at the University of Dundee, Scotland. He has been working in the field of accessibility and usability of technology for over 5 years, during which time he has advised many major educational, public and commercial organizations on web and software accessibility issues, and published widely in the academic and popular press on a variety of topics relating to widening access to digital media.

Preface

Welcome to volume 64 of the **Advances in Computers**. This series is the oldest continuing series, beginning in 1960, which describes the continuing evolution of the development of the computer and the field of information processing. Each year three volumes are produced presenting approximately 20 to 25 chapters that describe the latest technology and the use of computers today. This present volume, subtitled “Programming paradigms: The World Wide Web and other programming models,” is concerned about the changes in software development today. It is mostly about the impact that the Internet and the World Wide Web have on today’s use of computers, but includes other related development topics, such as the growing use of “open source” as a model for software development.

The Internet and the Web have had a profound effect on how information is obtained today. Billions of pages of data are freely available from computers worldwide. The problem is how to find that information. This is the role of the search engine, known locations where users can type in queries and obtain lists of pages that contain the information desired. However, how do search engines find their results? In the first chapter, “Automatic Evaluation of Web Search Services” by Abdur Chowdhury discusses the problem of “finding information needles in a worldwide haystack of search services.” He discusses various search strategies and various measures of determining how well the search worked.

Web search services are just one type of web service needed. As the commercial development of the web continues to grow, services are needed for handling financial transactions, inventory control, security of communications, and a whole host of other services needed to support a commercial enterprise. In Chapter 2, Sang Shin in “Web Services” discusses the various standards that are growing to handle this myriad of issues needing standardization. Without such standardization, it would be impossible for diverse industrial organizations to communicate effectively.

In Chapter 3, “A Protocol Layer Survey of Network Security” by John V. Harrison and Hal Berghel, expands upon the security standards of Chapter 2 and discusses how one violates the security of a system and can successful attack—or prevent an attack—on a computer system. It discusses the issues involved in intrusion detection systems and the various security defects, called vulnerabilities, which exist.

In Chapter 4, Roland T. Rust, P.K. Kannan and Anupama D. Ramachandran discuss “E-Service: The Revenue Expansion Path to E-Commerce Profitability.” Whereas Chapters 1 through 3 are concerned with a system builder’s view of certain web services, this chapter discusses the business view of the Web. How can one use the Web to develop a profitable electronic or e-business? “A customer-centric market view necessitates that firms learn the customer laws of the virtual world to retain and enhance their market share.”

Chapter 5 changes focus slightly and looks to the future. The World Wide Web is just the beginning of the integration of accessing multiple computers. In the future it is forecast that computers will be everywhere with large numbers of machines inter-linked with all aspects of your daily life. Originally called “ubiquitous computing” by Mark Weiser, it is now known as pervasive computing. Debashis Saha, in Chapter 5, discusses the implications of “Pervasive Computing: A Vision to Realize.” In his vision of pervasive computing, “the environment is saturated with a host of computing and communication capabilities which are gracefully integrated with daily life so that user will be able to interact with a smart environment from everywhere using a seemingly invisible infrastructure of various . . . devices.” What are the attributes of a pervasive computing environment, and what projects are now being undertaken to study this phenomenon?

“Open Source Software Development: Structural Tension in the American Experiment” by Coskun Bayrak and Chad Davis is the subject matter of Chapter 6. This brings up a new phenomenon in software development—open source. Traditionally, companies developed a product and then sold (or licensed) the software for users to run on their machine. Selling multiple copies of the software was how they acquired revenue. However, we now see another model for software development called open source. In this model, individuals contribute software to a central repository and anyone is able to download that repository, via the Web, and run the software for free. Systems such as the Linux operating system, the Eclipse development environment, the Apache web server, and the Mozilla web browser are all open source products. So the intriguing question is how do people who develop open source systems generate any revenue in order to stay in business? This chapter explores this question and other issues related to this concept.

The last chapter entitled “Disability and technology: Building barriers or creating opportunities?” by Peter Gregor, David Sloan and Alan F. Newell discusses an issue that cuts across all of the previous 6 chapters. It is generally assumed that individuals use computers by typing on a keyboard, moving a mouse, and reading a display screen. But what about those for whom these are difficult or impossible tasks? How does the blind interact with computers? The elderly often have problems with modern technology. What about other handicaps? This chapter discusses how one needs to develop computer systems that address the needs of these computer users as well. As

some of the earlier chapters clearly indicate, computers are becoming pervasive and without a degree of computer literacy, individuals will be cut off from many of the benefits of society. How do we provide those benefits to all?

I hope that you find these articles of interest. If you have any suggestions of topics for future chapters, or if you wish to contribute such a chapter, I can be reached at mvz@cs.umd.edu.

Marvin Zelkowitz
University of Maryland
College Park, MD, USA

Contents

CONTRIBUTORS	ix
PREFACE	xiii

Automatic Evaluation of Web Search Services

Abdur Chowdhury

0. Introduction	2
1. Relevance	4
2. A Brief History of Effectiveness Evaluations	5
3. Evaluation Metrics	9
4. Web Search Tasks	13
5. Estimating the Necessary Number of Queries	21
6. Automatic Web Search Evaluation	23
7. Automatic Evaluation Results	28
8. Intranet Site Search Evaluation	37
9. Conclusions and Future Work	39
Acknowledgements	40
References	40

Web Services

Sang Shin

1. Introduction on Web Services	46
2. Web Services Standards	58
3. Web Services Security	69
4. Identity Management Architecture	96
5. Business Web Services Standards: ebXML	102
6. Summary	106
References	106

A Protocol Layer Survey of Network Security

John V. Harrison and Hal Berghel

1. Introduction	110
2. Overview of TCP/IP	111
3. Offensive Techniques	115
4. Defenses	139
5. A Forecast of the Future	152
6. Defensive Precautions	153
7. Conclusion	156
References	156

E-Service: The Revenue Expansion Path to E-Commerce Profitability

Roland T. Rust, P.K. Kannan, and Anupama D. Ramachandran

1. Introduction	160
2. Building on Customer Equity	165
3. Customer Issues in E-Service	172
4. Serving the Customer	179
5. Marketing to Computers	185
6. Discussion and Conclusion	189
References	191

Pervasive Computing: A Vision to Realize

Debashis Saha

1. Introduction	197
2. Evolution of PerCom	201
3. PerCom Attributes	207
4. Functional Areas of PerCom	216
5. Harnessing Physical World	227
6. Major PerCom Projects	235
7. Conclusion	241
References	243

Open Source Software Development: *Structural Tension in the American Experiment*

Coskun Bayrak and Chad Davis

1. Introduction	248
2. Part I: The American Experiment	249
3. Part II: The Open Source Development Model	268
4. Open Societies: A Final Thought	278
References	281

Disability and Technology: Building Barriers or Creating Opportunities?

Peter Gregor, David Sloan, and Alan F. Newell

1. Introduction	284
2. Technology and the Digital Divide	286
3. Disabled People?	296
4. The Technical Benefits of Inclusive Design	309
5. Legislative Responsibilities	310
6. Accessible and Inclusive Design Practice	314
7. Support for Inclusive Design	321
8. Testing and Evaluation of Inclusive Design	328
9. Developments and Challenges	330
10. The Digital Divide Still Exists	342
References	343

AUTHOR INDEX	347
SUBJECT INDEX	353
CONTENTS OF VOLUMES IN THIS SERIES	369

Automatic Evaluation of Web Search Services

ABDUR CHOWDHURY

*Search & Navigation Group
America Online
USA
cabdur@aol.com*

Abstract

With the proliferation of online information, the task of finding information relevant to users' needs becomes more difficult. However, most users are only partly concerned with this growth. Rather, they are primarily focused on finding information in a manner and form that will help their immediate needs. It is essential to have effective online search services available in order to fulfill this need. The goal of this chapter is to provide a basic understanding of how to evaluate search engines' effectiveness and to present a new technique for automatic system evaluation.

In this chapter we explore four aspects of this growing problem of finding information needles in a worldwide haystack of search services. The first and most difficult is the exploration of the meaning of relevance to a user's need. The second aspect we examine is how systems have been manually evaluated in the past and reasons why these approaches are untenable. Third, we examine what metrics should be used to understand the effectiveness of information systems. Lastly, we examine a new evaluation methodology that uses data mining of query logs and directory taxonomies to evaluate systems without human assessors, producing rankings of system effectiveness that have a strong correlation to manual evaluations. This new automatic approach shows promise in greatly improving the speed and frequency with which these systems can be evaluated, thus, allowing scientists to evaluate new and existing retrieval algorithms as online content, queries, and the users' needs behind them change over time.

1. Introduction	2
2. Relevance	4
3. A Brief History of Effectiveness Evaluations	5
3.1. Cranfield 2 Experiments	5

3.2. TREC	7
4. Evaluation Metrics	9
4.1. Task Evaluation Metrics	9
5. Web Search Tasks	13
5.1. Manual Web Search Evaluations	16
5.2. The Changing Web	17
5.3. Changing Users' Interests	17
6. Estimating the Necessary Number of Queries	21
7. Automatic Web Search Evaluation	23
7.1. On-Line Taxonomies	25
7.2. Evaluation Methodologies	26
7.3. Engines Evaluated	27
8. Automatic Evaluation Results	28
8.1. Manual Evaluation	29
8.2. Taxonomy Bias	31
8.3. Automatic Evaluation	31
8.4. Stability	32
8.5. Category Matching	33
8.6. Automatic Effectiveness Analysis	36
9. Intranet Site Search Evaluation	37
10. Conclusions and Future Work	39
Acknowledgements	40
References	40

1. Introduction

The growth of information on the web has spurred much interest from both users and researchers. Users have been interested in the wealth of online information and services while researchers have been interested in everything from the sociological aspects to the graph theory involved in this hyperlinked information system. Because of the users' need to find information and services now available on the web, search engine usage is the second most common web activity after email. This fundamental need to find pertinent information has caused unprecedented growth in the market for both general and niche search engines. Google™, one of the largest web search engines, now boasts over 4 billion indexed HTML pages [24]. ResearchBuzz, a site that tracks search engine news, has reported some 30 new search products per month since the late 1990s [39]. These sites only begin to express the growth in available information and search engine activity that is being observed. With that growth, the basic research question we are interested in is: "How effective are these systems in finding relevant information?" This question is the focus of the chapter.

What does it mean to have an effective search service? There are many questions to consider when evaluating the effectiveness of a search service:

- Is the system responsive in terms of search time?
- Is the UI intuitive and well laid out?
- Is the content being searched both useful and complete?
- Does the search service help users fulfill their information need?
- Are the results presented with enough surrogate information for the users to understand whether their needs have been met?

These questions cover many aspects of a service's quality, from operational system characteristics [15], to the evaluation of the usability of the site [40]. Those issues are covered in other bodies of work and beyond the scope of this chapter. What is examined here is a service's ability to take an information need from a user and find the best set of results that satisfy that need. Additionally, we examine how a set of engines providing the same service can be examined and ranked in terms of how effectively they are meeting users' information request needs.

In Section 2 we explore the meaning of relevance, and ask the question "What is a good result from a search engine?" Since relevance is at the heart of information science, we present a brief background into prior efforts that attempt to provide a cogent definition of this elusive concept. In Section 3 we explore the history of search effectiveness evaluations, and the various aspects of effectiveness that must be studied. In Section 4 we explore the metrics used to understand these systems. In Section 5 we examine the web and the tasks users expect to accomplish when using web search services. In addition, we examine some of the factors that are specific to web systems in terms of changing user interests and content changes. We argue that because of constantly changing needs and content, traditional manual evaluations are not a tenable solution to understanding the effectiveness of these systems in any timely manner.

In Section 7 we examine a new methodology for automatically evaluating search services that is much less resource-intensive than human-reviewed relevance assessments. Performing human assessments on very large dynamic collections like the web is impractical, since manual review can typically only be done on a very small scale and is very expensive to repeat as content and users' needs change over time.

In Section 8 we examine automatic estimates of effectiveness on various tasks in relation to manual evaluations. In Section 9 we further explore how this approach can be applied to site and intranet search services. Lastly, in Section 10 we examine future research areas for these techniques.