

THE JOURNAL OF COMPUTERS

Volume 28
Number 1
January 1995



EDITED BY
ROBERT A. JOHNSON



Advances in **COMPUTERS**

Improving the Web

EDITED BY

MARVIN V. ZELKOWITZ

Department of Computer Science
University of Maryland
College Park, Maryland
USA

VOLUME 78



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier

32 Jamestown Road, London, NW1 7BY, UK

Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands

30 Corporate Drive, Suite 400, Burlington, MA 01803, USA

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

First edition 2010

Copyright © 2010 Elsevier Inc. All rights reserved

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher. Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at <http://elsevier.com/locate/permissions>, and selecting *Obtaining permission to use Elsevier material*

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

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-0-12-381019-9

ISSN: 0065-2458

For information on all Academic Press publications
visit our web site at elsevierdirect.com

Printed and bound in USA

10 11 12 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

Advances in
COMPUTERS
VOLUME 78

Contributors

Prof. Christian Breiteneder is a Full Professor for Interactive Systems with the Institute of Software Technology and Interactive Systems at the Vienna University of Technology. Dr. Breiteneder received the Diploma Engineer degree in Computer Science from the Johannes Kepler University in Linz in 1978 and a Ph.D. in Computer Science from the University of Technology in Vienna in 1991. Before joining the institute he was Associate Professor at the University of Vienna and had postdoc positions at the University of Geneva, Switzerland, and GMD (now Fraunhofer) in Birlinghoven, Germany. His current research interests include interactive media systems, media processing systems, content-based multimodal information retrieval, 3D user interaction, and augmented and mixed reality systems. His email address is breiteneder@ims.tuwien.ac.at.

Dr. Fabio Calefato received the M.Sc. and Ph.D. degrees in Computer Science from the University of Bari in 2002 and 2007, respectively. He is currently associated with the Collaborative Development Group at the University of Bari as postdoctoral research assistant. His main research interests focus on collaborative software engineering and computer-mediated communication theories and tools. He is also a member of the Eclipse Italian Community Group, since 2006, and AI*IA (Italian Association for Artificial Intelligence), since 2008. He can be contacted at calefato@di.uniba.it.

Prof. Wingyan Chung is an Assistant Professor in the Department of Operations and Management Information Systems in the Leavey School of Business at Santa Clara University. He received his Ph.D. in Management Information Systems from the University of Arizona and an M.S. and B.B.A. from the Chinese University of Hong Kong. His research interests include knowledge management, Web analysis and mining, data and text mining, business intelligence, information visualization, and human-computer interaction. He has over 50 refereed publications in *Journal of Management Information Systems*, *Communications of the ACM*, *IEEE Computer*, *International Journal of Human-Computer Studies*, *Decision Support Systems*, and

Journal of the American Society for Information Science and Technology, among others. A certified teacher since 1998, Dr. Chung has more than 10 years of teaching and curriculum development experiences in the United States and Hong Kong. He teaches information systems courses at both undergraduate and graduate levels. Contact him at wchung@scu.edu.

Prof. Filippo Lanubile is an Associate Professor of Computer Science at the University of Bari. From 1995 to 1997, he was a Research Associate in the Experimental Software Engineering Group at the University of Maryland. His research interests lie in the areas of software engineering and CSCW, focusing on social software and distributed software development. He is a recipient of a NASA Group Achievement Award (1996), an IBM Eclipse Innovation Award (2006), and an IBM Faculty Award (2008). He has been the Program Chair of the 3rd IEEE International Conference on Global Software Engineering (2008). He can be contacted at lanubile@di.uniba.it.

Prof. Ross A. Malaga is an Associate Professor of Management and Information Systems at the School of Business, Montclair State University. He received his Ph.D. from George Mason University. Dr. Malaga has published extensively in the area of electronic commerce in *Communications of the ACM*, *Electronic Commerce Research*, and the *Journal of Organization Computing and Electronic Commerce*. He serves on the editorial review boards of *Information Resources Management Journal* and the *Journal of Electronic Commerce in Organizations*. He can be reached at malagar@mail.montclair.edu.

Prof. Emilia Mendes is an Associate Professor in Computer Science at the University of Auckland in New Zealand. She has active research interests in the areas of empirical web and software engineering, evidence-based research, hypermedia, computer science and software engineering education, in which areas she has published widely and over 120 refereed publications, which include two books. Prof. Mendes is on the editorial board of the *International Journal of Web Engineering and Technology*, the *Journal of Web Engineering*, the *Journal of Software Measurement*, the *International Journal of Software Engineering and Its Applications*, the *Empirical Software Engineering Journal*, the *Advances in Software Engineering Journal*, the *Software Quality Journal*, and *Springer Journal of Internet Services and Applications*. She worked in the software industry for 10 years before obtaining a Ph.D. in 1999 in Computer Science from the University of Southampton (UK) and moving to Auckland (NZ). Her email address is emilia@cs.auckland.ac.nz.

Dalibor Mitrović is a Teaching and Research Associate with the Institute of Software Technology and Interactive Systems at the Vienna University of Technology. He received an M.Sc. degree in Computer Science in 2005 from the Vienna University of Technology, Austria. He pursues a Ph.D. in Computer Science focusing on multimodal information retrieval. His research interests include audio retrieval, real-time feature extraction, and computer vision. He can be reached at dalibor.mitrovic@computer.org.

Pedro Neves received his B.S. and M.S. degrees in Electronics and Telecommunications Engineering from the University of Aveiro, Portugal, in 2003 and 2006 respectively. From 2003 to 2006 he joined the Telecommunications Institute (IT), Aveiro, Portugal, and participated in the DAIDALOS-I and DAIDALOS-II (*Designing Advanced network Interfaces for the Delivery and Administration of Location independent, Optimized personal Services*) European-funded projects. Since 2006 he joined Portugal Telecom Inovação, Aveiro, Portugal, and he was involved in the WEIRD and, more recently, HURRICANE European-funded projects. Furthermore, since 2007 he is also pursuing a Ph.D. in Telecommunications and Informatics Engineering at the University of Aveiro. He has been involved in six book chapters, as well as more than 25 scientific papers in major journals and international conferences. His research interests are focused on broadband wireless access technologies, mobility and QoS management in *all-IP* heterogeneous environments, multicast and broadcast services, as well as mesh networks.

Dr. Kostas Pentikousis is a Senior Research Scientist at VTT Technical Research Centre of Finland, the foremost multidisciplinary applied research organization in Northern Europe. He holds a Ph.D. in Computer Science from the State University of New York at Stony Brook (2004). He received his B.Sc. and M.Sc. degrees in Computer Science from Aristotle University of Thessaloniki (1996; *summa cum laude*; ranked first) and State University of New York at Stony Brook (2000), respectively. Since 1996, he has been working in R&D positions in both industry and academia. He has been involved in several contract and joint research projects, including the EU-funded Ambient Networks (Phase 2), PHOENIX, WEIRD, and 4WARD; and the Future Internet program of the Finnish Strategic Centre for Science, Technology and Innovation in the field of ICT (TIVIT). Dr. Pentikousis has published more than 70 papers and book chapters in areas such as network architecture and design, mobile computing, applications and services, local and wide-area networks, and simulation and modeling, and has presented several tutorials on these topics. Dr. Pentikousis was an ERCIM Fellow in 2005 and is a member of IEEE, ACM, ICST, and TEK, the Finnish Association of Graduate Engineers.

Jarno Pinola received his M.Sc. degree in Telecommunications from the University of Oulu, Finland, in June 2008. He has worked at VTT Technical Research Centre of Finland since 2007 concentrating on wireless access technologies and mobility management. He has worked in both European and national projects covering these topics on different layers of the communication protocol stack. In the area of wireless communications, his current research interests include WiMAX system performance evaluations as well as LTE protocol studies. In addition, the energy-efficiency at the upper protocol levels in the next-generation cellular networks and wireless networking is part of his research agenda. He can be reached at jarno.pinola@vtt.fi.

Esa Piri received his M.Sc. from the University of Oulu, Finland, in Spring 2008. During his studies, he specialized in information networks and wrote his Master's Thesis on mobility management issues in heterogeneous networks. Currently, he is working as a Research Scientist in the field of seamless networking at VTT Technical Research Centre of Finland in Oulu, Finland. He can be contacted via e-mail at esa.piri@vtt.fi.

Dr. Susana Sargento joined the Department of Computer Science of the University of Porto in September 2002, and has been in the University of Aveiro and the Institute of Telecommunications since February 2004. She earned a Ph.D. in 2003 in Electrical Engineering. Dr. Sargento has been involved in several national and European projects, taking leaderships of several activities in the projects, such as the QoS and *ad hoc* networks integration activity in the FP6 IST-Daidalos Project. Her main research interests are in the areas of next generation and heterogeneous networks, infrastructure, mesh, and *ad hoc* networks, where she has published more than 150 scientific papers.

Matthias Zeppelzauer was born in Vienna, Austria, on October 15, 1980. He received the M.Sc. degree in Computer Science in 2005 from the Vienna University of Technology. Since 2006 Matthias Zeppelzauer works as a Ph.D. student at the Interactive Media Systems (IMS) group at the Vienna University of Technology. He has been employed in different research projects focusing on content-based audio and video retrieval. His research interests include content-based retrieval and recognition of sound, time series analysis, data mining, and multimodal media understanding. He can be reached at zeppelzauer@ims.tuwien.ac.at.

Preface

This is volume 78 in the series of *Advances in Computers* books. Currently, we publish three volumes, each containing 5 to 7 chapters, annually on new technology affecting the information technology industry. This series began in 1960 and is the oldest continuously published book series chronicling the ever changing computer landscape.

In this volume we again look at the World Wide Web. No longer an interesting research curiosity, the Web has become a dominant force in managing the flow of information internationally. It has become the primary source of information—both good and bad—for many researchers, it has become the backbone of multitrillion dollar (US) worldwide commerce, it has become the primary source of information between companies and governments to customers and the general public, and it is often the critical resource that manages the information flow within a company. In this volume we present six chapters that show how the Web is changing to address more complex issues in helping organizations manage their information technology resources.

In Chapter 1, “Search Engine Optimization—Black and White Hat Approaches,” Ross A. Malaga explores the role of search engines in managing the information flow through the Web. With the number of accessible web pages now being in the billions, one needs an index of the Web in order to access any desired page. Companies, such as Google, Yahoo, and Microsoft, are competing to be the search engine of choice. Each of these search engines try to present the user with the most relevant pages for any given query. On the other hand, companies want their own pages to be ranked high so that the search engines will display their pages to a given user’s query. Various techniques are employed by the search engines to rank pages and by companies to try and boost their own rankings. In this chapter, Dr. Malaga discusses this competing tension and describes the various approaches used to develop these rankings.

Chapter 2’s “Web Searching and Browsing: A Multilingual Perspective” by Wingyan Chung presents a different challenge presented by the emerging Web. As an outgrowth of the 1960s US Defense Advanced Research Projects Agency’s (DARPA) ARPANET, the Web had its roots in English text and the Roman

character set. But there has been significant growth in areas where other languages predominate, such as Japanese, Chinese, Arabic, among others. In this chapter, Dr. Chung discusses the issues in opening up the Web to other character sets and the issue of developing non-English search engines.

Dalibor Mitrović, Matthias Zeppelzauer, and Christian Breiteneder in Chapter 3's "Features for Content-Based Audio Retrieval" discuss still another attribute of the Web that has changed since its introduction. Initially, information was stored as textual documents, and search engines were needed to find appropriate documents containing appropriate textual information. But as the Web has grown, audio, video, graphics, and pictures are quickly overtaking the growth of text files. Tools such as iTunes for downloading music and video and sites such as YouTube.com store thousand of new audio and video files each day. Searching for the specific item is now based upon short textual tags placed with each such file. But there is increasing interest in searching such files by the features in the files themselves—how to locate features in a picture or specific sounds in an audio file. In this chapter, the authors discuss the searching of audio music files by the contents of the music itself.

Chapter 4 is related to the audio search issues in Chapter 3 but from a still different point of view. With the growth of the Web, more information is being processed and transmitted—not only text files. When the Web started to grow in the 1990s, you could say that the Web was a digital network that was being transmitted over analog telephone lines. However, somewhere around 10 years ago, the balance changed. Most telecommunications are now over digital lines, with telephony now being one of the technologies carried over this digital medium. The term "digital lines" is also becoming somewhat anachronistic with more and more communications over wireless radio links. Kostas Pentikousis, Jarno Pinola, Esa Piri, Pedro Neves, and Susana Sargento in "Multimedia Services over Wireless Metropolitan Area Networks" discuss these issues, including the current trend of using the Internet as a telephony network using "voice over IP" (VoIP) technologies.

With all of this development of new Web technologies, some of the criteria and assumptions (such as cost, schedules, reliability) concerning development of Web applications does not carry over from earlier "mainframe" development. Emilia Mendes looks at the software development problems of Web development in Chapter 5's "An Overview of Web Effort Estimation." How can we estimate the effort required to build new Web applications and how does this differ from older models? In this chapter, Dr. Mendes looks at various approaches to perform effort estimation for new Web applications.

In the last chapter, Fabio Calefato and Filippo Lanubile in "Communication Media Selection for Remote Interaction of *Ad Hoc* Groups" look at some of the implications of the technology described in the earlier chapters. With the increased speeds of the digital lines of the Web, with the increased capacity of the emerging

search engines, and with an increasing set of objects (e.g., text, pictures, video, telephony) being transmitted, one of the obvious impacts of these changes is that it is no longer necessary for groups to be working together in order to use and share such objects. Thus, we have an increasing dispersion of the workforce with social networks such as Facebook or Myspace providing mechanisms for communication over a distance. How can such groups communicate effectively? In this chapter, the authors discuss the various mechanisms that the Web provides for managing such interactions.

I hope that you find these chapters of interest. For me, volume 78 marks a significant milestone in my association with the *Advances in Computers* book series. The first volume I put together was volume 40 in 1995, which makes this the 39th volume produced under my direction. Or in other words, I have produced as many volumes as all previous series editors combined. I have found the process both interesting and rewarding, although at times a bit stressful in maintaining production schedules without sacrificing quality. I am always looking for new and interesting topics to appear in these pages. If you have any suggestions of topics for future chapters, or if you wish to contribute such a chapter yourself, I can be reached at mvz@cs.umd.edu.

Marvin Zelkowitz
College Park, Maryland

Contents

CONTRIBUTORS ix

PREFACE xiii

Search Engine Optimization—Black and White
Hat Approaches

Ross A. Malaga

1. Introduction 2

2. Background 4

3. The SEO Process 5

4. Black Hat SEO 18

5. Legal and Ethical Considerations 31

6. Conclusions 34

References 37

Web Searching and Browsing:
A Multilingual Perspective

Wingyan Chung

1. Introduction 42

2. Literature Review 43

3. A Multilingual Perspective 53

4. Case Studies 57

5. Summary and Future Directions 63
Acknowledgments 66
References 66

Features for Content-Based Audio Retrieval

**Dalibor Mitrović, Matthias Zeppelzauer,
and Christian Breiteneder**

1. Introduction 72
2. Background 74
3. Audio Feature Design 84
4. A Novel Taxonomy for Audio Features 94
5. Audio Features 99
6. Related Literature 133
7. Summary and Conclusions 139
Acknowledgments 139
References 139

**Multimedia Services over Wireless
Metropolitan Area Networks**

**Kostas Pentikousis, Jarno Pinola, Esa Piri, Pedro Neves,
and Susana Sargento**

1. Introduction 154
2. WiMAX Overview 157
3. Multimedia over WiMAX Reference Scenarios 170
4. Advances in Telephony and the Emergence of Voice over IP 178
5. VoIP over WiMAX 191
6. Remote Surveillance and IPTV over WiMAX 209
7. Summary and Outlook 216
Acknowledgments 219
References 219

An Overview of Web Effort Estimation

Emilia Mendes

| | |
|--|-----|
| 1. Introduction | 224 |
| 2. How to Measure a Technique's Prediction Accuracy? | 248 |
| 3. Which Effort Estimation Technique to Use? | 253 |
| 4. Web Effort Estimation Literature Survey | 254 |
| 5. Conclusions | 266 |
| References | 267 |

Communication Media Selection for Remote Interaction of *Ad Hoc* Groups

Fabio Calefato and Filippo Lanubile

| | |
|---|-----|
| 1. Introduction | 272 |
| 2. Task-Classification Frameworks | 277 |
| 3. Group Research | 281 |
| 4. CMC Theories | 285 |
| 5. Development of a Comprehensive Theoretical Framework | 300 |
| 6. Conclusions | 308 |
| Acknowledgments | 309 |
| References | 309 |

| | |
|--|-----|
| AUTHOR INDEX | 315 |
| SUBJECT INDEX | 327 |
| CONTENTS OF VOLUMES IN THIS SERIES | 339 |

Search Engine Optimization—Black and White Hat Approaches

ROSS A. MALAGA

Management and Information Systems, School of Business, Montclair State University, Montclair, New Jersey, USA

Abstract

Today the first stop for many people looking for information or to make a purchase online is one of the major search engines. So appearing toward the top of the search results has become increasingly important. Search engine optimization (SEO) is a process that manipulates Web site characteristics and incoming links to improve a site’s ranking in the search engines for particular search terms. This chapter provides a detailed discussion of the SEO process. SEO methods that stay within the guidelines laid out by the major search engines are generally termed “white hat,” while those that violate the guidelines are called “black hat.” Black hat sites may be penalized or banned by the search engines. However, many of the tools and techniques used by “black hat” optimizers may also be helpful in “white hat” SEO campaigns. Black hat SEO approaches are examined and compared with white hat methods.

| | |
|--|----|
| 1. Introduction | 2 |
| 2. Background | 4 |
| 2.1. Search Engines History and Current Statistics | 4 |
| 2.2. SEO Concepts | 5 |
| 3. The SEO Process | 5 |
| 3.1. Keyword Research | 5 |
| 3.2. Indexing | 9 |
| 3.3. On-Site Optimization | 11 |
| 3.4. Link Building | 15 |

- 4. Black Hat SEO 18
 - 4.1. Black Hat Indexing Methods 19
 - 4.2. On-Page Black Hat Techniques 19
 - 4.3. Cloaking 22
 - 4.4. Doorway Pages 23
 - 4.5. Content Generation 25
 - 4.6. Link Building Black Hat Techniques 25
 - 4.7. Negative SEO 30
- 5. Legal and Ethical Considerations 31
 - 5.1. Copyright Issues 31
 - 5.2. SEO Ethics 33
 - 5.3. Search Engine Legal and Ethical Considerations 33
- 6. Conclusions 34
 - 6.1. Conclusions for Site Owners and SEO Practitioners 35
 - 6.2. Future Research Directions 35
- References 37

1. Introduction

The past few years have seen a tremendous growth in the area of search engine marketing (SEM). SEM includes paid search engine advertising and search engine optimization (SEO). According to the Search Engine Marketing Professional Organization (SEMPO), search engine marketers spent over \$13.4 billion in 2008. In addition, this figure is expected to grow to over \$26 billion by 2013. Of the \$13.4 billion spent on SEM, about 10% (\$1.4 billion) was spent on SEO [1].

Paid advertising are the small, usually text-based, ads that appear alongside the query results on search engine sites (see Fig. 1). Paid search engine advertising usually works on a pay-per-click (PPC) basis. SEO is a process that seeks to achieve a high ranking in the search engine results for certain search words or phrases. The main difference between SEO and PPC is that with PPC, the merchant pays for every click. With SEO each click is free (but the Web site owner may pay a considerable amount to achieve the high ranking). In addition, recent research has shown that users trust the SEO (called organic) results and are more likely to purchase from them [2].

Industry research indicates that most search engine users only clicked on sites that appeared on the first page of the search results—basically the top 10 results. Very few users clicked beyond the third page of search results [3]. These results confirm the research conducted by Granka et al. [4], in which they found that almost 80% of

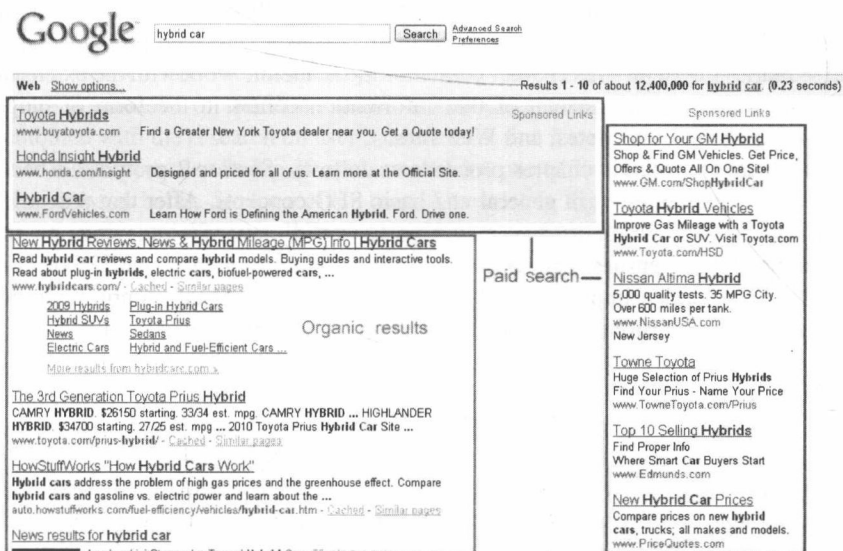


FIG. 1. Paid versus organic search results.

the clicks on a search engine results page came went to those sites listed in the first three spots.

SEO has become a very big business. Some of the top optimizers and SEO firms regularly charge \$20,000 or more per month for ongoing optimization. It is not uncommon for firms with large clients to charge them \$150,000 or more on a monthly basis [5].

Because of the importance of high search engine rankings and the profits involved, search engine optimizers look for tools, methods, and techniques that will help them achieve their goals. Some focus their efforts on methods aimed at fooling the search engines. These optimizers are considered "black hat," while those that closely follow the search engine guidelines would be considered "white hat." There are two main reasons why it is important to understand the methods employed by black hat optimizers. First, some black hats have proven successful in achieving high rankings. When these rankings are achieved, it means that white hat sites are pushed lower in the search results. However, in some cases these rankings might prove fleeting and there are mechanisms in place to report such sites to the search engines. Second, some of the tools and methods used by black hat optimizers can actually be used by white hat optimizers. In many cases, it is just a matter of scope and scale that separates black and white hat.