

WEB SEARCH: PUBLIC SEARCHING OF THE WEB

by

AMANDA SPINK

University of Pittsburgh, U.S.A.

and

BERNHARD J. JANSEN

The Pennsylvania State University, U.S.A.

KLUWER ACADEMIC PUBLISHERS

NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

Contents

Preface	vii
Foreword – <i>Tefko Saracevic</i>	xi
SECTION I: THE CONTEXT OF WEB SEARCH	
Chapter 1: Technological, Social and Organizational Context	3
Chapter 2: Human Information Behavior and Human Computer Interaction Context	19
Chapter 3: Research Design	35
SECTION II: HOW PEOPLE SEARCH THE WEB	
Chapter 4: Search Terms	55
Chapter 5: Search Queries	77
Chapter 6: Search Sessions	101
SECTION III: SUBJECTS OF WEB SEARCH	
Chapter 7: E-Commerce Web Searching	127
Chapter 8: Medical and Health Web Searching	137

Chapter 9: Sexually-Related Web Searching	149
Chapter 10: Multimedia Searching	161
SECTION IV: CONCLUSION	
Chapter 11: Key Findings, Trends, Further Research and Conclusions	181
Subject Index	191
Author Index	193

Preface

This book brings together results from the Web search studies we conducted from 1997 through 2004. The aim of our studies has been twofold: to examine how the public at large searches the Web and to highlight trends in public Web searching. The eight-year period from 1997 to 2004 saw the beginnings and maturity of public Web searching. Commercial Web search engines have come and gone, or endured, through the fall of the dot.com companies. We saw the rise and, in some cases, the demise of several high profile, publicly available Web search engines.

The study of the Web search is an exciting and important area of interdisciplinary research. Our book provides a valuable insight into the growth and development of human interaction with Web search engines. In this book, our focus is on the human aspect of the interaction between user and Web search engine. We do not investigate the Web search engines themselves or their constantly changing interfaces, algorithms and features. We focus on exploring the cognitive and user aspects of public Web searching in the aggregate. We use a variety of quantitative and qualitative methods within the overall methodology known as transaction log analysis.

Our studies examined large datasets of keywords, queries and search sessions provided by commercial Web search engine companies, but these companies provided limited or no access to the demographics of the individual or aggregate of Web search engines users. Our studies do not include analysis of query data from Web search engines where the data was unavailable to us. We are very grateful to commercial Web search engine companies who were generous enough to provide large query data sets for academic analysis, including Excite.com, AskJeeves.com, AlltheWeb.com

and AltaVista.com. Their support for our research is an outstanding example of the beneficial cooperation that can occur between industry and academia.

The authors have contrasting backgrounds. This contrast contributed greatly to the interdisciplinary nature of our studies. The first author (Spink) is an information scientist who has worked with, taught and researched human interaction with information systems and search engines since 1980. The second author (Jansen) is a computer scientist who has worked with, taught and researched information systems, information retrieval, Web content design, and search engines since 1986. The authors started working jointly on researching public Web searching when Excite.com offered a large dataset of Web queries to attendees at the 1997 Association of Computing Machinery Special Interest Group on Information Retrieval Conference (ACM-SIGIR) in Philadelphia.

PURPOSE AND APPROACH

Our book has three objectives. First, we provide an overview and synthesis of Web searching research within a broad theoretical framework. Second, we offer in-depth analysis of Web searching within several topical domains. Third, we provide an overview and synthesis of our research findings on public Web searching and highlight trends in public Web searching.

The book is organized in four sections to reflect these objectives:

Section I: The Context of Web Search

Section II: How People Search the Web

Section III: Subjects of Web Search

Section IV: Trends and Future Directions

Thus, Section I describes the broad framework and context for research examining public Web searching, including the technological, social, organizational, human information behavior and human computer interaction levels of analysis. This section also describes the research designs used in our studies. Section II focuses on providing an overview and synthesis of public Web searching. In particular we focus on users' search terms and topics, querying and search sessions. Section III discusses the major subjects of Web searching in four areas: e-commerce, medical health, sexual search, and multimedia search. Section IV discusses the key findings of our studies, the trends in public Web search we observed, including the growth of more complex search behaviors, and future research directions.

Each section contains one or more chapters relating to the broader topic area of the section. Each chapter is stand-alone, in respect of presenting a complete picture of the particular topic, including its own reference list. The chapters are also cross-referenced where appropriate to illustrate how that topic meshes with the broader area of Web search.

AUDIENCE

This book is a valuable resource for Web searching researchers, educators, and practitioners. The primary audience is researchers and students in the fields of information science, computer science, information systems, cognitive science, and related disciplines. The book is a valuable research resource for those investigating Web searching. This book is an appropriate text for undergraduate, graduate and doctoral level courses in areas of information and Web retrieval, online information management, information science, human information behavior, digital libraries, Web content structuring, and management information systems.

Web consultants, search engine optimization specialists, Web masters, providers of online content and services, and those working within Web search engine companies would find this book useful for research-based insights into how people seek and search for electronic information. In addition, anyone who searches the Web will find the book a fascinating and enlightening read.

ACKNOWLEDGMENTS

A good portion of this book is a result of many previous studies conducted by the authors, often with other researchers and students. We would like to thank Tefko Saracevic from Rutgers University for helping to launch this major research project, and Dietmar Wolfram from the University of Wisconsin-Milwaukee who contributed to the data analysis and key publications.

We also thank Jan Pedersen from Alta Vista, Per Gunan from AlltheWeb.com, and Jack Xu and Doug Cutting from Excite@home.com, for generously providing the large datasets of Web queries that formed the initial basis of our research. We also acknowledge our colleagues Seda and Huseyin Cenk Ozmutlu from Uludag University (Turkey) who analyzed some of the transaction logs and co-authored publications.

We also thank our students and other researchers who collaborated on the query analysis and co-authored publications, including Mohammad Abbas,

Judy Bateman, Carol Chang, Abby Goodrum, Alice Goz, Okan Guner, Shaoyi He, Ali Hurson, Daniel Lorence, Stephanie Milchak, Pirkko Nyganen, Minsoo Park, Anthony Pfaff, Megan Pomeroy, Andrew Koricich, Michelle Sollenberger, and Yin Yang. We would also like to acknowledge the following people for their reviews of our book chapters – James Wang, Magy Sief El-Nasr, Dongwon Lee, Praseniit Mitra, Karen Jansen and Peter Spink.

Amanda Spink

School of Information Sciences
University of Pittsburgh

Bernard J. Jansen

School of Information Sciences and Technology
The Pennsylvania State University