

Search User Interfaces

This book focuses on the human users of search engines and the tool they use to interact with it: the search user interface. The truly worldwide reach of the Web has brought with it a new realization among computer scientists and laypersons of the enormous importance of usability and user interface design. In the last ten years, much has become understood about what works in search interfaces from a usability perspective, and what does not. Researchers and practitioners have developed a wide range of innovative interface ideas, but only the most broadly acceptable make their way into major Web search engines. This book summarizes these developments, presenting the state of the art of search interface design, both in academic research and in deployment in commercial systems.

Many books describe the algorithms behind search engines and information retrieval systems, but the unique focus of this book is specifically on the user interface. The intended audiences are industry professionals who are designing systems that use search interfaces as well as graduate students and academic researchers who investigate information systems.

Dr. Marti A. Hearst is a Professor in the School of Information at the University of California, Berkeley. She received B.A., M.S., and Ph.D. degrees in computer science from University of California Berkeley and was a member of the research staff of Xerox PARC from 1994 to 1997. Her research focuses on search user interfaces, computational linguistics, and information visualization. She is an Okawa Foundation Fellow and an IBM faculty Fellow and is a recipient of the NSF CAREER Grant, a Google faculty research grant, and more than \$2 million in government research funding. She has advised and consulted for numerous search-related companies.

Cambridge University Press
978-0-521-11379-3 - Search User Interfaces
Marti A. Hearst
Frontmatter
[More information](#)

Search User Interfaces

Marti A. Hearst

University of California, Berkeley



Cambridge University Press
978-0-521-11379-3 - Search User Interfaces
Marti A. Hearst
Frontmatter
[More information](#)

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA
www.cambridge.org
Information on this title: www.cambridge.org/9780521113793

© Marti A. Hearst 2009

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2009

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Hearst, Marti.

Search user interfaces / Marti A. Hearst.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-521-11379-3

1. Web search engines. 2. User interfaces (Computer systems)

3. Human-computer interaction. I. Title.

TK5105.884.H43 2009

006.7'784-dc22

2009025735

ISBN 978-0-521-11379-3 Hardback

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party Internet Web sites referred to in
this publication and does not guarantee that any content on such Web sites is,
or will remain, accurate or appropriate.

Cambridge University Press
978-0-521-11379-3 - Search User Interfaces
Marti A. Hearst
Frontmatter
[More information](#)

In memory of my mother, Jean.

Contents

<i>Preface</i>	<i>page xi</i>
1 <i>The Design of Search User Interfaces</i>	1
1.1 Keeping the Interface Simple	1
1.2 A Historical Shift in Search Interface Design	3
1.3 The Process of Search Interface Design	5
1.4 Design Guidelines for Search Interfaces	6
1.5 Offer Efficient and Informative Feedback	7
1.6 Balance User Control with Automated Actions	14
1.7 Reduce Short-Term Memory Load	18
1.8 Provide Shortcuts	22
1.9 Reduce Errors	23
1.10 Recognize the Importance of Small Details	25
1.11 Recognize the Importance of Aesthetics in Design	26
1.12 Conclusions	28
2 <i>The Evaluation of Search User Interfaces</i>	29
2.1 Standard Information Retrieval Evaluation	30
2.2 Informal Usability Testing	34
2.3 Formal Studies and Controlled Experiments	36
2.4 Longitudinal Studies	46
2.5 Analyzing Search Engine Server Logs	47
2.6 Large-Scale Log-Based Usability Testing (Bucket Testing)	49
2.7 Special Concerns with Evaluating Search Interfaces	52
2.8 Conclusions	63

3 Models of the Information Seeking Process	64
3.1 The Standard Model of Information Seeking	64
3.2 Cognitive Models of Information Seeking	66
3.3 The Dynamic (Berry-Picking) Model	67
3.4 Information Seeking in Stages	69
3.5 Information Seeking as a Strategic Process	71
3.6 Sensemaking: Search as Part of a Larger Process	80
3.7 Information Needs and Query Intent	82
3.8 Conclusions	90
4 Query Specification	91
4.1 Textual Query Specification	91
4.2 Query Specification via Entry Form Interfaces	101
4.3 Dynamic Term Suggestions During Query Specification	105
4.4 Query Specification Using Boolean and Other Operators	107
4.5 Query Specification Using Command Languages	114
4.6 Conclusions	118
5 Presentation of Search Results	120
5.1 Document Surrogates	120
5.2 KWIC, or Query-Oriented Summaries	122
5.3 Highlighting Query Terms	128
5.4 Addition Features of Results Listings	130
5.5 The Effects of Search Results Ordering	135
5.6 Visualization of Search Results	139
5.7 Conclusions	139
6 Query Reformulation	141
6.1 The Need for Reformulation	141
6.2 Spelling Suggestions and Corrections	142
6.3 Automated Term Suggestions	144
6.4 Suggesting Popular Destinations	151
6.5 Relevance Feedback	152
6.6 Showing Related Articles (More Like This)	154
6.7 Conclusions	156
7 Supporting the Search Process	157
7.1 Starting Points for Search	157
7.2 Supporting Search History	162

<i>Contents</i>	ix
7.3 Supporting the Search Process as a Whole	164
7.4 Integrating Search with Sensemaking	168
7.5 Conclusions	173
8 <i>Integrating Navigation with Search</i>	174
8.1 Categories for Navigating and Narrowing	175
8.2 Categories for Grouping Search Results	177
8.3 Categories for Sorting and Filtering Search Results	180
8.4 Organizing Search Results via Table-of-Contents Views	182
8.5 The Decline of Hierarchical Navigation of Web Content	187
8.6 Faceted Navigation	188
8.7 Navigating via Social Tagging and Social Bookmarking	196
8.8 Clustering in Search Interfaces	199
8.9 Clusters vs. Categories in Search Interfaces	208
8.10 Conclusions	210
9 <i>Personalization in Search</i>	211
9.1 Personalization Based on Explicit Preferences	213
9.2 Personalization Based on Implicit Relevance Cues	221
9.3 Combining Implicit and Explicit Information	229
9.4 Searching over Personal Information	231
9.5 Conclusions	232
10 <i>Information Visualization for Search Interfaces</i>	234
10.1 Principles of Information Visualization	235
10.2 Techniques for Interactive Visualization	237
10.3 The Effects of Data Types on Information Visualization	238
10.4 The Difficulties with Visualizing Nominal Data	240
10.5 Visualization for Query Specification	247
10.6 Visualizing Query Terms Within a Large Document	252
10.7 Visualizing Query Terms Within Retrieval Results	254
10.8 Visualizing Faceted Navigation	267
10.9 Visualizing Search Results as Clusters and “Starfields”	273
10.10 3D Visualization in Search	278
10.11 Conclusions	280
11 <i>Information Visualization for Text Analysis</i>	281
11.1 Visualization for Text Mining	281
11.2 Visualizing Document Concordances and Word Frequencies	286

x	<i>Contents</i>
11.3 Visualizing Literature and Citation Relationships	294
11.4 Conclusions	296
12 <i>Emerging Trends in Search Interfaces</i>	297
12.1 Mobile Search Interfaces	297
12.2 Multimedia Search Interfaces	306
12.3 Social Search	317
12.4 A Hybrid of Command and Natural Language Search	322
12.5 Conclusions	323
<i>Appendix: Additional Copyright Notices</i>	325
<i>Bibliography</i>	329
<i>Index</i>	365
<i>Author Index</i>	375

Preface

Search is an integral part of peoples' online lives; people turn to search engines for help with a wide range of needs and desires, from satisfying idle curiosity to finding life-saving health remedies, from learning about medieval art history to finding video game solutions and pop music lyrics. Web search engines are now the second most frequently used online computer application, after email. Not long ago, most software applications did not contain a search module. Today, search is fully integrated into operating systems and is viewed as an essential part of most information systems.

Many books on information retrieval describe the *algorithms* behind search engines and information retrieval systems. By contrast, this book focuses on the human users of search systems and the tool they use to interact with them: the search *user interface*. Because of their global reach, search user interfaces must be understandable by and appealing to a wide variety of people of all ages, cultures, and backgrounds, and for an enormous variety of information needs.

The truly worldwide reach of the Web has brought with it a new realization among computer scientists and laypersons alike of the enormous importance of usability and user interface design. In the last ten years, much has become understood about what works in search interfaces from a usability perspective, and what does not. Researchers and practitioners have developed a wide range of innovative interface ideas, but only the most broadly acceptable make their way into major Web search engines. This book attempts to summarize these developments, presenting the state of the art of search interface design, both in academic research and in deployment in commercial systems.

This is a fast-changing field, and any attempt to summarize the state of the art will no doubt soon be proven obsolete. Nonetheless, certain principles and techniques seem to hold steady over the years, and there is much that is now known about search interfaces that should stand for at least the near future.

Book Overview

This book outlines the human side of the information seeking process and focuses on the aspects of this process that can best be supported by the user interface. It describes the methods behind user interface design generally, and search interface design in particular, with an emphasis on how best to evaluate search interfaces. It discusses research results and current practices surrounding user interfaces for query specification, display of retrieval results, grouping retrieval results, navigation of information collections, query reformulation, search personalization, and the broader tasks of sensemaking and text analysis. Much of the discussion pertains to Web search engines, but the book also covers the special considerations surrounding search of other information collections. The chapters are elaborated on in the following paragraphs.

The Design of Search User Interfaces (Chapter 1) introduces the ideas and practices surrounding user interface design generally, and search interface design in particular. It opens with an analysis of why Web search interfaces appear standardized and relatively simple compared to other interfaces and places modern search interfaces into a historical context. The remainder of the chapter is a summary of interface design guidelines as applied specifically to search interfaces. This chapter is intended to be useful for those people who do not have time to read the entire book but want to understand best practices and problems to avoid in the design of search user interfaces.

The Evaluation of Search User Interfaces (Chapter 2) is a companion to the chapter on design, as user-centered design requires tight-coupling with evaluation. The chapter summarizes the key methods for evaluating search interfaces: informal studies, formal studies, field studies, longitudinal studies, and large-scale log-based studies (also known as bucket testing). This is followed by advice about best practices and special considerations to keep in mind when evaluating search interfaces.

Models of the Information Seeking Process (Chapter 3) summarizes the theoretical models that have been proposed about how people seek information. These models are the foundation on which much of search interface design is based. These include the standard model, the cognitive model, the dynamic (berry-picking) model, information seeking as a strategic process (including cost-structure analysis and foraging theory), orienteering and incremental strategies, and the theory of sensemaking. This is followed by a discussion of information needs and query intent, including attempts to create taxonomies of searcher's information needs and intents, by manually and automatically analyzing queries and online behavior.

Query Specification (Chapter 4) is the first of a set of three chapters that describe interfaces to support the interlocked information seeking cycle of query specification, viewing of retrieval results, and query reformulation. This chapter summarizes both research and the state of current practice in search interface design for query specification, including textual queries, natural language questions, query specification forms, dynamic feedback, queries using Boolean and other operators, faceted queries, and command-based queries.

Presentation of Search Results (Chapter 5) is the second of three chapters on interface support for the standard information seeking cycle. This chapter summarizes research as well as the state of current practice for displaying search results pages. Topics include document surrogates, properties of results listings, summaries or extracts as used in search results, and user response to search results ordering.

Query Reformulation (Chapter 6) is the third of three chapters on interface support for the standard information seeking cycle. This chapter discusses the need for and frequency of query reformulation, followed by interface ideas that support reformulation from both research and the current state of practice. Specifically, these are spelling suggestions and corrections, automated suggestions for query refinement and expansion, suggesting popular destinations, relevance feedback, and suggesting related articles.

Supporting the Search Process (Chapter 7) is a capstone to the previous three chapters, describing interfaces that encompass and augment the full standard process of information seeking. Topics include interfaces to

support finding starting points for search, using history and re-finding, and to support the sensemaking process that often accompanies but is broader than search.

Integrating Navigation with Search (Chapter 8) discusses interfaces to support the integration of browsing of information structures with directed search, primarily in information collections (as opposed to for the Web as a whole). Topics include using categories to sort, filter, and group search results; organizing results by table-of-contents-like views; faceted navigation; and automatically derived clusters for organizing search results. The chapter concludes with a discussion of the tradeoffs of using categories versus clusters for search results organization.

Personalization in Search (Chapter 9) explores the emerging area of using information about individual users to influence search results ordering, to create automated alert services, and to tailor information recommendations to users. There is intense interest in research and industry surrounding information personalization, although most attempts to personalize information are still short of their mark.

Information Visualization for Search Interfaces (Chapter 10) is the first of two closely related chapters on the use of information visualization in search and text analysis. This first chapter provides a brief introduction to the main principles and techniques used in visualization of abstract information (as opposed to scientific visualization, which renders real-world objects in visual form). It also discusses why visualization of nominal data – of which text is composed – is difficult to do effectively. It then describes some of the many attempts to use information visualization to improve query specification and display of retrieval results, as well as to give overviews of information collections. Unfortunately, in most cases, usability studies incorporating these visualizations find that in the best case they do not improve peoples' performance, and in the worse case they slow people down or cause them to make errors. That said, in many cases study participants find visualizations to be appealing, at least at first exposure if not for extended use.

Information Visualization for Text Analysis (Chapter 11) describes information visualization for text analysis, which seems to be a more successful application area for visualization of textual information. Although primarily of interest for analysts and specialists, as opposed to

for everyday search use, these techniques are often creative in design and captivating to view.

Emerging Trends in Search Interfaces (Chapter 12) closes the book with a discussion of areas of search that are still relatively new but promise to be of increasing importance in the coming years. Topics include mobile search, multimedia search, social search, and a hybrid of command-based and natural language search.

There are a number of topics related to search that this book does not cover. These include interfaces for database systems, Search Engine Optimization (SEO), the role of advertising in search, spam detection and elimination, and ranking algorithms. This book assumes that the reader is familiar with the technical basics behind search engines and information retrieval, including crawling, indexing, ranking, Boolean queries, and PageRank. (For those who are not, see the Related Books section for suggested readings.)

This book is an update and expansion of a chapter written in 1998 for the book *Modern Information Retrieval*, Baeza-Yates and Ribeiro-Neto (Eds.), Addison Wesley, 1999. At that time, little was known definitively about which ideas result in usable search interfaces. In the intervening ten years, much has been learned. Thus, one goal of this book is to back up every statement with verification from the literature. This can be challenging when much of the knowledge is locked up in industry, but fortunately, a number of recent papers have appeared that share insights from the major Web search engines. (An exception is made for Chapter 1, which is intended to be a summary of best practices encapsulated into one chapter; unsupported statements made there are verified in later chapters.) Although a large proportion of the references are necessarily drawn from research from the last few years, readers are also exposed to early foundations and ideas.

Using This Book

This book has two intended primary audiences. The first is academic researchers, graduate students, and those teaching graduate-level courses in information retrieval, user interfaces, and other information management-related topics. The second intended audience is practitioners who design and build search interfaces. Although the book makes heavy use of academic references, an attempt has been made to keep the

language and concepts approachable. Instructors may want to view this book as having two main parts, with Chapters 1–7 covering search interface fundamentals and Chapters 8–12 covering advanced topics.

The contents of this book are available online at <http://searchuserinterfaces.com>. Updates to the subject matter presented in this book will appear at the Web site.

Related Books

For a nice introduction to the mathematical foundations and algorithms for search, geared primarily towards undergraduates, see *Introduction to Information Retrieval*, by Manning, Rhagavan, and Schütze, Cambridge University Press, 2008. For a more advanced research-oriented book on a wider range of topics related to search, see *Modern Information Retrieval, 2nd Edition*, by Baeza-Yates and Ribeiro-Neto (Eds.), Addison Wesley, 2009 (to appear).

For details on Web search algorithms, see *Mining the Web: Analysis of Hypertext and Semi Structured Data* by Chakrabarti, Morgan Kaufmann, 2002, and for details on link-based algorithms as well as general Web search algorithms, see *Google's PageRank and Beyond: The Science of Search Engine Rankings* by Langville and Meyer, Princeton University Press, 2006. For implementing search engines, see *Managing Gigabytes* by Witten, Moffat, and Bell, Morgan Kauffman, 1999, and *Lucene in Action* by Gospodnetic and Hatcher, Manning Publications, 2004.

To date, there is no other academic book that focuses on search user interfaces. The most closely related is *Information Seeking in Electronic Environments* by Marchionini, Cambridge University Press, 1995, which focuses on the search process rather than on interfaces for search. *Finding Out About: A Cognitive Perspective on Search Engine Technology and the WWW* by Rik Belew, Cambridge University Press, 2008, is a new edition of a book first published in 2000 and describes basic algorithms as well as some cognitive properties of search.

Books written by and for practioners include *Information Architecture for the World Wide Web, 3rd Edition* by Morville and Rosenfeld, O'Reilly Media, 2006, which describes design of information architecture, including two chapters on search, and *Designing Web Navigation, Optimizing the User Experience*, by Kalback and Gufstafson, O'Reilly Media, 2007, which discusses navigation design for Web sites.

A Note on Terminology

The words *user*, *searcher*, and *information seeker* are used interchangeably in this book to indicate a hypothetical or an actual person using a search system. Some authors object to the word *user*, both because they feel it reduces a person to what they are doing with a computer and because of its association with recreational drug use. Its use is, however, standard in the field and is convenient to write with, and so this book is a user user. By contrast, the word *participant* is used to refer to a person who voluntarily participates in a usability study.

This book also unapologetically uses the pronoun *they* and the possessive *their* to refer to the third person singular as a way to avoid making explicit (and unnecessarily distracting) gender distinctions.

Disclaimer

The author has been employed by, consulted for, and/or received research gifts or grants from the following institutions whose ideas, products, or projects are mentioned in this book: AltaVista, DeepDyve, Google, IBM, Microsoft, Powerset, (Xerox) PARC, University of California Berkeley, Yahoo, and Zvents. No compensation has been received or is expected in exchange for mentioning these organizations' products or ideas in this book.

Acknowledgments

I wish to thank the following people for commenting on this manuscript: Bob Glushko made extensive comments on an early draft that led to my writing a single chapter that can be read in isolation (the design chapter). I am grateful to Ben Shneiderman, who was generous with an extensive conversation about the visualization chapters. Daniel Russell provided invaluable feedback on the design, query specification, and results presentation chapters, and Dan Rose and Anne Aula provided detailed comments on the evaluation chapter. I am also grateful to Omar Alonso, Anne Aula, Stephen Few, Greg Linden, Gary Marchionini, Avi Rappoport, and Jamie Teevan for comments on other chapters.

I would like to thank the hundreds of former master's students who have taken my courses in User Interface Design, Information

Visualization, and Information Organization and Retrieval. The more than 50 projects I oversaw in the User Interface Design course were invaluable for deepening my understanding of how the interface design process unfolds, and the pitfalls as well as the successful paths toward good design. I also thank the former master's students, PhD students, and post-docs who worked with me on search interface research projects: Anna Divoli, Ame Elliott, Jennifer English, Melody Ivory, Kevin Li, Preslav Nakov, Ariel Schwartz, Rashmi Sinha, Emilia Stoica, Kirsten Swearingen, Michael Wooldridge, and Ka-Ping Yee.

I would like to thank the University of California, Berkeley, for granting me a sabbatical that allowed me to finish the writing of this book, Lauren Cowles at Cambridge University Press for acting as my editor and liaison to the publisher, and Ricardo Baeza-Yates and Berthier Ribiero-Neto, whose request for a chapter for the revision of *Modern Information Retrieval* led me to write an entire book.

Closer to home, I thank my brother Ed for persistently encouraging me to write a book, my sister Dor for inspiring me by finishing her first novel this year, my parents for surrounding me with books and a love of words and science, and Emmi for frequent play breaks. Finally, I thank Carl for being there throughout, in the best way possible.