
▶ **Going Beyond Google**

**The Invisible Web in
Learning and Teaching**

Jane Devine

Francine Egger-Sider

Neal-Schuman Publishers, Inc.
New York London

Published by Neal-Schuman Publishers, Inc.
100 William St., Suite 2004
New York, NY 10038

Copyright © 2009 Neal-Schuman Publishers, Inc.

All rights reserved. Reproduction of this book, in whole or in part, without written permission of the publisher, is prohibited.

Printed and bound in the United States of America.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992.

Library of Congress Cataloging-in-Publication Data

Devine, Jane, 1947-

Going beyond Google : the Invisible Web in learning and teaching / Jane Devine,
Francine Egger-Sider.

p. cm.

Includes bibliographical references and index.

ISBN 978-1-55570-633-3 (alk. paper)

1. Invisible Web. 2. Invisible Web—Study and teaching. 3. Internet searching. 4. Database searching. I. Egger-Sider, Francine. II. Title.

ZA4237.D4 2009

025.042—dc22

2008050850

► Contents

List of Figures and Tables	vii
Preface	ix
Acknowledgments	xii

► PART I
UNDERSTANDING THE DIVISION BETWEEN
THE VISIBLE AND INVISIBLE WEB

1. Characteristics of the Invisible Web	3
Introduction	3
Relationship between Search Engines and the Invisible Web	5
Nature of the Relationship	5
A Constantly Changing Relationship	8
Implications of the Relationship	8
Characteristics That Make the Invisible Web Important	9
Nature of the Material Included	9
Databases	9
Deep Web Sites	11
Other Invisible Web Resources	12
Searchability of the Invisible Web	13
Complexity	13
Tools	13
Navigation Skills	14
Lack of Stability	14
Place in the Research World	14
Teachers	15
Librarians	15
Researchers in Any Professional Field	15
Dissatisfied Web Searcher	15
Anyone Wanting to Excel	15
Conclusions	16
References	16
2. Use of the Web for Research	17
Introduction	17
Statistics of Reliance on the Visible Web	17

User Studies	18
Nationwide Studies of College Students and Adults	19
Nationwide Studies of Middle School and High School Students	23
Studies of Individual Institutions	24
Other Research About Student Information-seeking Behavior	26
Faculty Research Patterns	26
Summary	28
Problems Associated with Using General-purpose Search	
Engines Exclusively	29
Retrieval Limits	29
Search Strategy Problems	30
Evaluation Issues	31
Conclusions	32
References	32

▶ PART II
FINDING AND UTILIZING THE CONTENTS
OF THE INVISIBLE WEB

3. Introducing Students to the Invisible Web	37
Introduction	37
Prerequisites for Learning about the Invisible Web	37
The Information Need	37
Student Attitudes and Preparedness	38
When and Where to Teach Use of the Invisible Web	39
Stages for Learning about the Invisible Web	40
Stage One Concepts: Web Searching Basics	41
Stage Two Concepts: Presenting the Invisible Web	47
Stage Three Concepts: The Riches of the Invisible Web	52
Opportunities for Invisible Web Instruction	53
Introducing the Invisible Web in Reference Work	53
Introducing the Invisible Web in the One-shot Instruction Period	54
Introducing the Invisible Web in Credit Courses	55
Introducing the Invisible Web with Course Management Systems	56
Other Opportunities for Introducing the Invisible Web	57
Summary	57
References	57
4. Further Exploration of the Invisible Web	59
Introduction	59
Activity 1: To Know	60
Objective	60
Setup	60
Conclusion	65

Activity 2: To Access	65
Objective	65
Setup	65
Conclusion	70
Activity 3: To Evaluate	70
Objective	70
Setup	70
Conclusion	73
Conclusions	74
References	74
5. Internet Research Strategies: An Example	75
Introduction	75
Using Invisible Web Content in Research: Case Study	75
Search Engine Results	77
Library Database Results	80
Subject Directory Results	81
Narrowing the Search	83
Conclusions	86
References	87
Appendix: Search on “Microfinance” in INFOMIME	88
6. A Sampler of Tools for Mining the Invisible Web	93
Introduction	93
General Subject Directories	93
Librarians’ Internet Index (www.lii.org)	93
The Open Directory Project (www.dmoz.org)	95
INFOMINE: Scholarly Internet Resource Collections (http://infomine.ucr.edu)	95
Other Subject Directories of Interest	96
Databases	97
THOMAS (http://thomas.loc.gov/)	97
LexisNexis Academic (www.lexisnexis.com)	97
ERIC—Education Resources Information Center (www.eric.ed.gov/)	99
OAIster.org (www.oaister.org)	100
Specialized Search Engines	101
Invisible Web Search Engines	102
Vertical Search Engines	102
Niche Search Engines	105
Semantic Web Search Engines	106
Conclusions	107
References	107

▶ PART III
NARROWING THE GAP BETWEEN THE
VISIBLE AND INVISIBLE WEB

7. Visible versus Invisible Web: Shifting Boundaries	111
Introduction	111
Relative Size of the Surface Web and the Invisible Web	112
Commercial General-purpose Search Engine Inroads into the Invisible Web	113
Google Scholar	114
Google Book Search	114
Crawling through HTML Forms	115
Yahoo! Initiatives	115
OCLC	116
Deep Web Search Engines	116
Vertical Search Engines	117
Microsoft	117
ProQuest DigitalCommons@	118
Web 2.0	118
Initiatives to Narrow the Gap between Visible and Invisible Web	118
Next-generation Library Catalogs	119
Federated Searching	121
Indexing the Invisible Web	122
Other Noncommercial Initiatives	122
Conclusions	124
References	124

▶ APPENDICES

Appendix A. Selected Additional Readings	129
Appendix B. Tools for Teaching the Invisible Web	133
Appendix C. ACRL Information Literacy Standards, Performance Indicators, and Outcomes	139
Index	147
About the Authors	156

► List of Figures and Tables

FIGURES

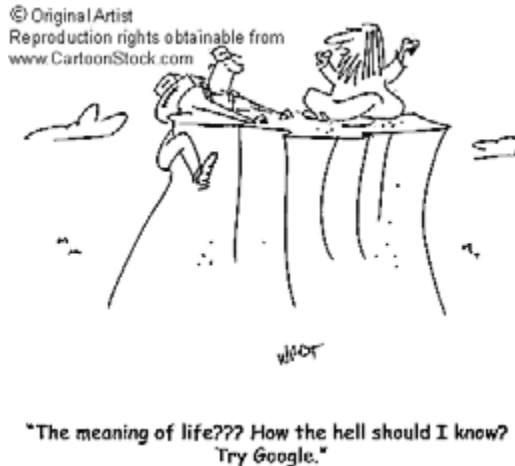
Figure 1-1	Search Engines Dragging a Net Across the Web's Surface	7
Figure 1-2	Harvesting the Deep and Surface Web	7
Figure 4-1	INFOMINE Results Mapped	72
Figure 4-2	Google Results Mapped	73
Figure 5-1	First Ten Results of Google Search on "Microfinance"	77
Figure 5-2	First Ten Results of Search for "Microfinance" in Academic Search Premier	81
Figure 5-3	Results of Google Search for "Microfinance and Poverty"	84
Figure 5-4	First Eight Results of Search for "Microfinance and Poverty" in EBSCOhost's Academic Search Premier	85
Figure 5-5	Comparing Search Results Using Three Interfaces	87
Figure 6-1	Librarian's Internet Index (www.lii.org)	94
Figure 6-2	Open Directory Project (www.dmoz.org)	95
Figure 6-3	INFOMINE (http://infomine.ucr.edu)	96
Figure 6-4	THOMAS (http://thomas.loc.gov/)	98
Figure 6-5	LexisNexis Academic (www.lexisnexis.com)	98
Figure 6-6	ERIC (www.eric.ed.gov/)	99
Figure 6-7	OAIster (www.oaister.org)	100
Figure 6-8	Horizontal, Vertical, and Niche Tools Relationship	101
Figure 6-9	Scirus (www.scirus.com)	103
Figure 6-10	Science.gov (www.science.gov)	104
Figure 6-11	ScienceResearch.com (www.scienceresearch.com)	105
Figure 6-12	hakia (www.hakia.com)	106
Figure 7-1	How Much Information 2003	112
Figure 7-2	Illustration of the World Wide Web Content Layers—Summer 2005	113

TABLES

Table 2-1	Results of OCLC and Pew Reports User Studies	21
Table 2-2	Tools Used for Research Tasks	22
Table 3-1	Stage One Concepts Mapped to Information Literacy Standards	43
Table 3-2	Stage Two Concepts Mapped to Information Literacy Standards	49

Table 4-1	Google Alerts for “Deep Web,” “Invisible Web,” and “Hidden Web” for January and February 2007	62
Table 4-2	Comparison of Yahoo! and Google Linking	69
Table 5-1	Description of Web Sites and Checklist Criteria Applied	78
Table 5-2	Analysis of Web Sites and Checklist Criteria Applied	82
Table 6-1	Summary of Internet Research Tools	107

► Preface



Is the meaning of life available through Google? So it would seem from the information-seeking behavior of many students, especially at the college level. In a recent cartoon by Chris Wildt, a guru on a mountaintop sends a petitioner to Google to learn the meaning of life. Wildt pokes fun at the underlying assumption that everything is on the Web and that Google can find it all. In fact, researchers who use only one general-purpose search engine, such as Google or Yahoo!, to find information on the Web are like investors who accept a return of \$25.00 as full payment on a loan of \$100.00. Even the best search engine offers at most only 25 percent of the information available on the Web.

A vast wealth of information that is not accessible via Google and its analogs is indeed available through the Web. Yet, this vast array of knowledge remains a secret to most students and researchers because general-purpose search engines do not index them. Thus, they fall into what we call the "Invisible Web." The authors (and most librarians) advocate teaching students about the limitations of general-purpose search engines and introducing them to the riches available through other interfaces. Students should receive a fuller picture of the information world in which general-purpose search engines inhabit a small part. They can be encouraged to explore further and to understand the many information choices open to them.

Going Beyond Google: The Invisible Web in Learning and Teaching proposes to address the dilemma created by the explosion of online information available at anyone's fingertips. This book argues that it is up to faculty, teachers, and librarians to fundamentally change "the research process" as currently undertaken by

students and redirect learners toward information beyond that found through general-purpose search engines.

It is true that students find one-stop shopping for sources through Google, or other general-purpose search engines, much easier than having to learn and navigate the interfaces of several different publishers. At the same time, commercial firms are also making inroads into the Invisible Web by enlarging the pool of resources indexable by general-purpose search engines. The gap between the visible and the Invisible Web is definitely narrowing. Still, relying exclusively on general-purpose search engines gives the student the false impression that he or she has access to whatever is available on the Internet. It is imperative to educate students to the complexities of research.

Anyone interested in the concept of the Invisible Web can profit from this book. However, the book is specifically geared toward the teaching faculty at every level of the educational spectrum, including librarians who can mold a student's research habits. The first step is to have students attain an awareness of the division between the "surface" or "visible" Web and the "deep" or "Invisible" Web. The next step is to help students find and utilize the contents of the Invisible Web. This objective can be accomplished in various ways, either by requiring an analysis of the sources used or by requiring the use of Invisible Web content for a particular research assignment. The authors present a number of activities that can be undertaken with a class, irrespective of the discipline taught.

To set the foundation for the main argument, Chapter 1, "Characteristics of the Invisible Web," examines its relationship to the surface Web, the nature of the material of which it is comprised, and its searchability. The last section addresses the various constituencies that could benefit from mining the Invisible Web.

Because a discussion of the Invisible Web in learning and teaching cannot be undertaken without an understanding of how users, primarily students, use the Web for research and why they overwhelmingly start, and often end, their research on the free Web, Chapter 2, "Use of the Web for Research," reviews the major user behavior studies undertaken in the past ten years. It is important to identify which features of general-purpose search engines are so attractive to all users, and to students in particular, that they use these search engines to the exclusion of many other sources. Such knowledge can empower teachers and librarians to intervene in the research process to redirect the choice of resources to be used.

Chapter 3, "Introducing Students to the Invisible Web," is a staged introduction to concepts about the Invisible Web that readers can adapt and use in the classroom or the library, irrespective of discipline. Each concept is mapped to the information literacy competency standards of the Association of College and Research Libraries and the American Association of School Librarians. The idea of the Invisible Web can be woven into reference work, library instruction, regular class instruction, and course management software, such as Blackboard.

Chapter 4, “Further Exploration of the Invisible Web,” offers structured exercises for more in-depth understanding of the Invisible Web. They can be used by anyone or introduced in a classroom setting.

Chapter 5, “Internet Research Strategies,” provides a detailed analysis of a research assignment on microfinance given to a hypothetical second-year student in a college international studies class. The search is run through three search tools: Google; EBSCOhost’s Academic Search Premier, a subscription database offered in most college libraries; and INFOMINE, an online subject directory. The results of the three searches are analyzed and evaluated. This type of comparison between the results of two or three different sources can be used in any class where research is required. It forces students to look at the results of their searches and evaluate the validity of the sources found and their relevance to the assignment at hand.

Chapter 6, “A Sampler of Tools for Mining the Invisible Web,” examines the different tools, such as directories, databases, and search engines, that teachers and their students can use to explore the Invisible Web. A chart recommending when to use these tools is included.

Finally, Chapter 7, “Visible versus Invisible Web,” analyzes the shifting boundaries between the “visible” or “surface” Web and the “deep” or “Invisible” Web. As general-purpose search engines uncover more and more sources by indexing them through crawlers, the border between the coexisting realms of “visible” and “Invisible” are constantly shifting. In addition, new technological innovations are mushrooming, whether in the commercial sector, with new initiatives such as Google Scholar and Google Book Search, or in the library world, with the advent of new library online catalogs such as the one at North Carolina State University Libraries, running on Endeca software. All of these developments have the beneficial result of bringing more content into the visible Web, allowing researchers easier access to quality sources.

Last, appendices include selected additional readings, teaching tools that can be used to impart the concept of the Invisible Web, including graphics, and the text of the *Information Literacy Competency Standards for Higher Education* of the Association of College and Research Libraries.

The authors hope *Going Beyond Google: The Invisible Web in Learning and Teaching* will expand the borders of information beyond resources identified by Google search results and help open a wider world of primary and secondary sources that can better inform learners and expand the horizons of knowledge.

► Acknowledgments

This book would not have been possible without the support of several people whom the authors would like to take this opportunity to thank. First, we want to express our gratitude to Professor Louise Fluk who was tireless about reading and rereading our chapters and whose invaluable advice was always sought. Thanks must also go to Professor Steven Ovadia who would always drop everything to help with technical advice and know-how. Thanks also to Nasrin Rahman for her great patience with the authors' preoccupation with "the book." Finally, a loud thank you must go to our respective families who lived this experience with us. We will share every success with them.
