

/THEORY/IN/PRACTICE

The Art of SEO

Mastering Search Engine Optimization



Eric Enge
Stephan Spencer
Rand Fishkin
Jessie C. Stricchiola
Foreword by John Battelle

O'REILLY®

Praise for **The Art of SEO**

“Roll up your sleeves, buckle your seat belt, and take your foot off the brake. You are about to go on a journey from the very basics to the very high-end, enterprise level, and then into the future of the art of SEO. These four authors have been involved in Internet marketing from the very start and have hands-on experience. These are not pundits in search of an audience but practitioners who have actually done the work, know how it’s done, and have the scars to prove it. This is a dynamite primer for the beginner and a valued resource for the expert. Clear, concise, and to the point, it may not make you laugh or make you cry, but it will make you smart and make you successful.”

—JIM STERNE, PRODUCER OF THE EMETRICS MARKETING OPTIMIZATION SUMMIT (WWW.EMETRICS.ORG) AND CHAIRMAN OF THE WEB ANALYTICS ASSOCIATION (WWW.WEBANALYTICSASSOCIATION.ORG)

“In your hands is a definitive collection of SEO knowledge from four leading practitioners of the art. This book is required reading for my company, and we’re also recommending it to our clients.”

—ADAM AUDETTE, PRESIDENT, AUDETTEMEDIA, INC., AND LEAD SEO FOR ZAPPOS.COM

“*The Art of SEO* will give you the opportunity to learn from all the people who taught me what I know today. Whether you are a beginner or a seasoned veteran, this is the most informative resource on the search marketing industry. An extremely important read for the ever-growing and fast-paced online marketing world.”

—TONY ADAM, SEO PROGRAM MANAGER AT YAHOO!

“If you have a business online, you should be looking for ways to develop traffic from search engines. That being the goal, your next stop should be a quiet corner, a tea or coffee, and a fresh copy of *The Art of SEO*. Seriously, this book can shave years off the learning curve for anyone thinking of delving into the world of search marketing. *The Art of SEO* walks you through the most important steps in planning and executing a top-flight program. Easy to understand and well written, this book walks you through everything you need to understand to be successful with your own SEO campaigns. The authors of this book are trusted individuals whose repeated, proven success working with SEO marks them as leaders in the field. Read now, prosper now.”

—DUANE FORRESTER, AUTHOR OF *HOW TO MAKE MONEY WITH YOUR BLOG*, SENIOR PROGRAM MANAGER (SEO), AND MICROSOFT SEMPO (SEARCH ENGINE MARKETING PROFESSIONAL ORGANIZATION) BOARD MEMBER

“In an increasingly search-driven world, where the majority of all search engine traffic is driven by natural listings, implementing an effective SEO program is critical for leading brands. *The Art of SEO* is essential reading for advertisers who want to understand effective SEO strategy and execution to drive consumer engagement and response. This book will help advertisers win through SEO.”

—NICK BEIL, CHIEF EXECUTIVE OFFICER, PERFORMICS

“*The Art of War* isn’t about Chinese pottery, and *The Art of SEO* isn’t a paint-by-numbers kit. This 500-page book is a comprehensive guide to search engine optimization strategies and tactics written by four SEO experts: Eric Enge, Stephan Spencer, Rand Fishkin, and Jessie Stricchiola. The chapters on keyword research, developing an SEO-friendly website, creating link-worthy content and link marketing, and tracking results and measuring success are must-reads for anyone interested in mastering search engine optimization.”

—GREG JARBOE, PRESIDENT OF SEO-PR, AND AUTHOR OF *YOUTUBE AND VIDEO MARKETING: AN HOUR A DAY*

“In *The Art of SEO*, industry luminaries Stephan Spencer, Rand Fishkin, Eric Enge, and Jessie Stricchiola successfully translate their deep collective knowledge into a straightforward and engaging guide covering it all—fundamentals, advanced techniques, management strategies, and an array of useful tools and tips. It’s required reading for anyone looking to maximize search engine traffic to their site.”

—MARK KAUFMAN, ASSOCIATE VICE PRESIDENT, CNET FINDABILITY, CBS INTERACTIVE

“Collectively, Rand, Eric, Jessie, and Stephan know more about SEO than anyone else on the planet. You want to master SEO? Listen to this dream team!”

—AVINASH KAUSHIK, AUTHOR OF *WEB ANALYTICS 2.0* AND *WEB ANALYTICS: AN HOUR A DAY*

“Watching a movie with my 16-year-old, we saw a detective in a library searching through Microfiche! (My daughter had no clue.) This was a sharp reminder of how far search has come in such a short period. No book has yet been written that so completely educates web marketers on this critical and rapidly advancing tool.”

—CHRIS BAGGOTT, CO-FOUNDER AND CEO, COMPENDIUM SOFTWARE, CO-FOUNDER, EXACTTARGET

“You may know enough about search engine optimization to be dangerous, but *The Art of SEO* will make you formidable.”

—CHRIS PIRILLO, INTERNET ENTREPRENEUR, *CHRIS.PIRILLO.COM*

“This is a valuable and comprehensive treatment of fundamental and complex SEO topics. The authors intelligently cover an impressive array of diverse tactics and essential skills that range from practical to strategic. A section on testing provides process examples for research and analysis, an often-ignored requirement in an industry where the target is always moving. *The Art of SEO* is the kind of book that ends up highlighted, dog-eared, and coffee-stained.”

—ALEX BENNERT, CHIEF SEARCH STRATEGIST, *WALL STREET JOURNAL*

“*The Art of SEO* is the perfect complement to the science of conversion optimization. This book is a must-read volume by four highly regarded industry veterans.”

—BRYAN EISENBERG, NEW YORK TIMES BEST-SELLING AUTHOR OF *CALL TO ACTION AND ALWAYS BE TESTING*

“This book covers everything an SEO needs to know. Not only a great introduction to the field, *The Art of SEO* encompasses the most current and most advanced strategies and tactics to deliver measurable results. Perhaps of even greater value, the reader is sure to glean insights into launching highly profitable SEO campaigns today and well into the future.”

—RICHARD ZWICKY, FOUNDER AND PRESIDENT, *ENQUISITE*

“Businesses online face unprecedented competition for the time and dollars of consumers. The authors have captured their deep knowledge of patterns and best practices in search, and made it accessible to anyone with a stake in driving traffic and bottom-line results. This book is packed with information, yet still an easy read. It will take the mystery out of search engine marketing and empower you to build a successful business online.”

—JEREMIAH ANDRICK, ONLINE CUSTOMER ACQUISITION MANAGER, LOGITECH,
AND FORMER PROGRAM MANAGER FOR MICROSOFT BING

“An essential guide to best practices and cutting-edge tactics that belongs on the desk of all search marketing professionals.”

—CHRIS SHERMAN, EXECUTIVE EDITOR, *SEARCH ENGINE LAND*

"The Art of SEO is really about the science of SEO. This detailed and practical guide to SEO mastery, from a panel of all-star practitioners, will give you the edge. Get it before your competitors do!"

—**TIM ASH, CEO OF SITE TUNERS.COM, AND AUTHOR OF THE BEST-SELLING
*LANDING PAGE OPTIMIZATION: THE DEFINITIVE GUIDE TO TESTING AND
TUNING FOR CONVERSIONS***

"There are no better names in the search marketing industry to write a book on the art of SEO than these four authors. Each author has gems of knowledge to share individually, and all of them teaming up to create a single book is like discovering a treasure."

—**BARRY SCHWARTZ, NEWS EDITOR, SEARCH ENGINE LAND, AND EDITOR,
SEARCH ENGINE ROUNDTABLE**

"I've known and followed the authors for years, each a bona fide expert in their own right. Their collective wisdom in one book is truly a gift to the industry. This has to be the #1'must-read' SEO book ever written."

—**ERIC WARD, A.K.A. LINK MOSES**

"This book dashed any thought I had of writing my own SEO book. The funny thing is, I am ecstatic because it is everything I could have hoped for—and more—for anyone interested in SEO. First, it was written by a veritable dream team of SEOs who are known as the best in class, and second, it is jam-packed with not just the golden rules of SEO but also some great tips and techniques that are sure to make this a desk-side reference for everyone."

—**ROSS DUNN, CEO/FOUNDER, STEP FORTH WEB MARKETING, INC.**

"The Art of SEO is a masterpiece in search engine optimization techniques. Whether you're technical or creative, whether a coder, a designer, a copywriter, or a PR professional, you need this book."

—**ANDY BEAL, CO-AUTHOR OF RADICALLY TRANSPARENT, FOUNDER OF
TRACKUR, AND FOUNDER OF MARKETING PILGRIM**

"The utmost compliments go to the team that pulled together *The Art of SEO*. As a professional educator, I can attest to the extreme difficulty of making SEO understandable and interesting. This is a *must-read* for every entrepreneur, marketer, and Internet professional to understand the fundamentals and importance of SEO to their business."

—**AARON KAHLOW, FOUNDER, ONLINE MARKETING SUMMIT**

The Art of SEO moves deftly from the search engine basics to the latest in cutting-edge tactics. The authors have distilled the breathtaking complexity of search ranking factors and SEO workflow into immediately usable information for practitioners. Until now, there has been no single definitive guide to recommend to newcomers in the field. Finally, it's here."

—ANDREW GOODMAN, FOUNDER AND PRINCIPAL, PAGE ZERO MEDIA

"As a co-author of a book people refer to as the 'Bible of Search Marketing,' you might think that I wouldn't recommend other search books. Not so. But I recommend only excellent search books written by outstanding search experts. *The Art of SEO* easily clears that high standard and is a must-read for anyone serious about organic search success."

—MIKE MORAN, CO-AUTHOR OF *SEARCH ENGINE MARKETING, INC.*, AND
AUTHOR OF *DO IT WRONG QUICKLY*

"More than just a 'how to do SEO' book, *The Art of SEO* gives even the uninitiated a sense of the overall search industry, its technological importance, and an insider's view of how to strategically leverage search for marketing success. That being said, this is not a fluffy marketing book—it's a practical, tactical manual that should be required reading for anyone responsible for their company's web reputation. I was particularly impressed at how the book helps readers think more about valuing and leveraging a wide range of business assets that companies have already invested in and probably don't even realize it—and how to fully empower these assets to drive engagement and ultimately revenue. There's a reason why SEO remains a top priority for today's most successful companies."

—DANA TODD, CMO OF NEWSFORCE, AND PAST CHAIRPERSON OF SEMPO
(SEARCH ENGINE MARKETING PROFESSIONAL ORGANIZATION)

The Art of SEO provides the nuts and bolts of SEO and beyond. This book gives you the tools you need to grok and apply a wide range of strategies immediately—giving you the plans to build, and to remodel when necessary—and it assists with hammering and painting, too. SEO is more than just keywords, copy, and layout. The authors deftly guide you through the constantly evolving search engine landscape, in all its aspects. Does SEO permeate throughout everything you publish online? It should. Make each page, each word, each link count. It doesn't matter whether your site is for lead generation, sales, or reputation building. Every web master or marketer needs a copy of this book on the shelf—or a stack of them to distribute to their team."

—KELLY GOTO, FOUNDER AND CEO, GOTOMEDIA

"Written by in-the-trenches practitioners, *The Art of SEO* is a well-written step-by-step guide providing sensible and practical advice on how to implement a successful SEO program. The authors have created a readable and straightforward guide filled with concise and easily adopted strategies and tactics any online business can use. I now have a great resource to recommend when people ask, 'Know any good books on SEO?'"

—DEBRA MASTALER, PRESIDENT, ALLIANCE-LINK AND THE LINK SPIEL BLOG

"An amazingly well-researched, comprehensive, and authoritative guide to SEO from some of the most well-respected experts in the industry...highly recommended for anyone involved in online marketing."

—BEN JESSON, CEO AND CO-FOUNDER OF CONVERSION RATE EXPERTS

"In *The Art of SEO*, Eric Enge, Stephan Spencer, Rand Fishkin, and Jessie Stricchiola have taken on the daunting task of compiling a comprehensive, step-by-step walk-through of what it takes to rank well on search. They go well beyond the usual tactical aspects, addressing fundamental challenges like understanding user intent, integrating an SEO culture within your organization, properly measuring success, and managing an SEO project. This is a deep, deep dive into the world of organic optimization, and you couldn't ask for better guides than Spencer, Fishkin, Enge, and Stricchiola. Clear a place on your desk, because this is going to be your SEO bible."

—GORD HOTCHKISS, PRESIDENT, ENQUIRO SEARCH SOLUTIONS

"*The Art of SEO* will delight and inform the next generation of skilled SEO practitioners. The author lineup is the search industry equivalent of the classic musical supergroup—think the Crosby, Stills, Nash, Young of search."

—DAVID SZETELA, FOUNDER AND CEO, CLIX MARKETING

"The disciplined and scientific practice of natural search engine optimization is critical to brand awareness and new customer acquisition. *The Art of SEO* has transformed what has historically been a misunderstood and mystical marketing strategy into an easy-to-comprehend, actionable guide to understanding and navigating the inner and outer workings of SEO."

—SETH BESMERTNIK, CEO AND CO-FOUNDER, CONDUCTOR

"If you do a search in Google for 'search engine optimization,' 'SEO,' or any similar term, you will find countless outdated articles that promote practices that are not very useful these days—website submissions, link exchanges, altering meta keyword tags, etc.

"These seemingly useful tactics do very little for the ultimate goal of an effective SEO campaign—to drive meaningful traffic. Because search engines have changed significantly in the last 10 years, many of these practices are no longer necessary, while some, like massive link exchanges, are actually considered search engine 'spam.' If these are what you are expecting from *The Art of SEO*, you will be positively disappointed. Sure, this book is about everything you will ever need to know about SEO now and in the near future, but after my personal technical review of all its merits, I can guarantee you that I couldn't find a single piece of nonsensical advice. If you only want one book, get this one. You can start from zero and become a SEO master in record time."

—HAMLET BATISTA, PRESIDENT AND CEO, RANKSENSE

"In *The Art of SEO*, these four industry leaders have left no stone unturned in their quest to deliver one of the ultimate resources on search engine optimization that has ever been published."

—CHRIS WINFIELD, PRESIDENT AND CO-FOUNDER, 10E20

"DO NOT BUY THIS BOOK. Please. I beg of you. If you compete with us or any of our clients, do *not* buy this book. It's become our go-to source for anything—and everything—we need to know about successful search engine optimization."

—AMY AFRICA, CEO, EIGHT BY EIGHT

"*The Art of SEO* is an extraordinary book chock-full of the most current wisdom on the constantly changing world of search. This is the one resource you need to help you build a strong foundation for success with SEO."

—DANA VANDEN HEUVEL, FOUNDER AND CHIEF THOUGHT LEADER AT
MARKETINGSAVANT

"There are no better guides through the world of SEO—the combined experience of these authors is unparalleled. I can't recommend highly enough that you buy this book."

—WILL CRITCHLOW, CO-FOUNDER, DISTILLED

"Regardless of whether you're a beginner or an expert search marketer, *The Art of SEO* delivers! From keyword research and search analytics to SEO tools and more!"

—KEN JURINA, PRESIDENT AND CEO, EPIAR

"Rarely does a work so thoroughly deconstruct the art and science of SEO: what it is, how it works, who makes it happen, and why it is important to the modern firm."

—SARA HOLOUBEK, CONSULTANT AND COLUMNIST, AND SEMPO PRESIDENT
THROUGH APRIL 2010

"This must-have book by industry heavyweights is a milestone. The material is convincing and compelling. Most important of all, the ideas make powerful strategies for successfully marketing sites online."

—DISA JOHNSON, CEO, SEARCHRETURN

"Finally, a guide to the perplexing world of SEO by some of its most accomplished practitioners. *The Art of SEO* has become my bible of search. Full of clear examples, cutting-edge research, and smart marketing strategies, this is a fun read that can help get your site the search ranking it deserves."

—HOWIE JACOBSON, AUTHOR OF *ADWORDS FOR DUMMIES*

"Fantastic read! This is a *must-read* for anyone in our industry. This book is a veritable textbook, and almost certainly will become part of any curriculum on the subject."

—JEFF QUIPP, CEO, SEARCH ENGINE PEOPLE

"There is an art (and science) to search engine optimization. It's not always easy, it's not always obvious, and the results depend a lot on what the major search engines are tinkering with under their own hoods. Thankfully, there is a book like *The Art of SEO* to shine a light, give you some clues, and help you get ahead of your competitors."

—MITCH JOEL, PRESIDENT OF TWIST IMAGE, AND AUTHOR OF *SIX PIXELS OF SEPARATION*

“Search engine optimization continues to evolve in exciting ways. This book provides one of the most comprehensive guides to planning and executing a full SEO strategy for any website. It will be an important reference for SEO professionals, business owners, and anyone who wants to succeed in the SEO field.”

—KHALID SALEH, PRESIDENT, INVESP

“Presenting the inner mechanics of search engine optimization is a daunting task, and this book has accomplished it with flair. The book reveals the closely guarded secrets of optimizing websites in a straightforward, easy-to-understand format. If you ever wanted to unravel the mysteries of the most enigmatic discipline on the Internet, this is the book you want as your guide. This book is so comprehensive and well written, it just might put me out of a job.”

—CHRISTINE CHURCHILL, PRESIDENT, KEYRELEVANCE

“Search is becoming increasingly important for marketers to understand. Businesses that don’t include organic search as a core component to their marketing efforts risk being left behind as their competitors and their customers move forward. This book covers a lot of ground—it’s full of stats, advice, and tools to equip marketers in their organic search efforts.”

—VANESSA FOX, AUTHOR OF *MARKETING IN THE AGE OF GOOGLE*

“This book has it all—SEO tactics, history, and articles to back up their writing (which are great for forwarding via email), combined with how to create and execute an SEO program. It’s the ultimate reference guide for any SEO.”

—JESSICA BOWMAN, SEOINHOUSE.COM

The Art of SEO



Eric Enge, Stephan Spencer, Rand Fishkin, and Jessie C. Stricchiola
foreword by John Battelle

O'REILLY®
Beijing • Cambridge • Farnham • Köln • Sebastopol • Taipei • Tokyo

The Art of SEO

by Eric Enge, Stephan Spencer, Rand Fishkin, and Jessie C. Stricchiola

Copyright © 2010 Eric Enge, Stephan Spencer, Rand Fishkin, and Jessie Stricchiola. All rights reserved.
Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (<http://my.safaribooksonline.com>). For more information, contact our corporate/institutional sales department: 800-998-9938 or corporate@oreilly.com.

Editor: Mary E. Treseler

Production Editor: Lorannah Dimant

Copyeditor: Audrey Doyle

Proofreader: Katie Nopper DePasquale

Indexer: Margaret Troutman Zaig

Cover Designer: Mark Paglietti

Interior Designer: David Futato

Illustrator: Robert Romano

Printing History:

October 2009: First Edition.

Nutshell Handbook, the Nutshell Handbook logo, and the O'Reilly logo are registered trademarks of O'Reilly Media, Inc. *The Art of SEO*, the image of a booted racket-tail hummingbird, and related trade dress are trademarks of O'Reilly Media, Inc.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and O'Reilly Media, Inc. was aware of a trademark claim, the designations have been printed in caps or initial caps.

While every precaution has been taken in the preparation of this book, the publisher and authors assume no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.



This book uses RepKover™, a durable and flexible lay-flat binding.

ISBN: 978-0-596-51886-8

[M]

1255027183

I'd like to dedicate this book to Beth, Rob, Valerie, and Kristian, who are, without question, the principal joys of my life. I'd also like to thank the countless people in the SEO community who have helped me along the way.

—Eric Enge

I dedicate this book to my beautiful daughters, Chloë, Ilsa, and Cassandra, for their love and support, and for tolerating my workaholic tendencies. They are wise beyond their years. They keep me grounded.

—Stephan Spencer

I'd like to dedicate this book to the SEO community and note that, like Einstein, if I have had any success, it's because I've stood on the shoulders of giants—to all those who practice, evangelize, and support SEO, thank you and keep up the great work. I'm also immensely grateful to Geraldine DeRuiter, love of my life, and the most talented writer this side of Hemingway.

—Rand Fishkin

I would like to dedicate this book to my late aunt, Anna Oliynik, who gave her tremendous love, support, and encouragement during the writing process, and who was taken from us before her time during the writing of this book. May your spirit be flying with the hummingbirds. I would also like to give my grateful thanks to Rosemary Garrison and to my mother, Robin, for all their love and support.

—Jessie C. Stricchiola

C O N T E N T S

FOREWORD	xix
PREFACE	xxi
1 THE SEARCH ENGINES: REFLECTING CONSCIOUSNESS AND CONNECTING COMMERCE	1
<i>The Mission of Search Engines</i>	2
<i>The Market Share of Search Engines</i>	2
<i>The Human Goals of Searching</i>	2
<i>Determining Searcher Intent: A Challenge for Both Marketers and Search Engines</i>	5
<i>How People Search</i>	9
<i>How Search Engines Drive Commerce on the Web</i>	14
<i>Eye Tracking: How Users Scan Results Pages</i>	14
<i>Click Tracking: How Users Click on Results, Natural Versus Paid</i>	16
<i>Conclusion</i>	22
2 SEARCH ENGINE BASICS	23
<i>Understanding Search Engine Results</i>	24
<i>Algorithm-Based Ranking Systems: Crawling, Indexing, and Ranking</i>	30
<i>Determining Searcher Intent and Delivering Relevant, Fresh Content</i>	41
<i>Analyzing Ranking Factors</i>	49
<i>Using Advanced Search Techniques</i>	53
<i>Vertical Search Engines</i>	65
<i>Country-Specific Search Engines</i>	74
<i>Conclusion</i>	78
3 DETERMINING YOUR SEO OBJECTIVES AND DEFINING YOUR SITE'S AUDIENCE	79
<i>Setting SEO Goals and Objectives</i>	79
<i>Developing an SEO Plan Prior to Site Development</i>	84
<i>Understanding Your Audience and Finding Your Niche</i>	85
<i>SEO for Raw Traffic</i>	91
<i>SEO for E-Commerce Sales</i>	91
<i>SEO for Mindshare/Branding</i>	92
<i>SEO for Lead Generation and Direct Marketing</i>	93
<i>SEO for Reputation Management</i>	93
<i>SEO for Ideological Influence</i>	95
<i>Conclusion</i>	96
4 FIRST STAGES OF SEO	97
<i>The Major Elements of Planning</i>	97
<i>Identifying the Site Development Process and Players</i>	100
<i>Defining Your Site's Information Architecture</i>	101

<i>Auditing an Existing Site to Identify SEO Problems</i>	106
<i>Identifying Current Server Statistics Software and Gaining Access</i>	116
<i>Determining Top Competitors</i>	118
<i>Assessing Historical Progress</i>	122
<i>Benchmarking Current Indexing Status</i>	124
<i>Benchmarking Current Rankings</i>	126
<i>Benchmarking Current Traffic Sources and Volume</i>	127
<i>Leveraging Business Assets for SEO</i>	131
<i>Combining Business Assets and Historical Data to Conduct SEO/Website SWOT Analysis</i>	132
<i>Conclusion</i>	134
5 KEYWORD RESEARCH	135
<i>The Theory Behind Keyword Research</i>	135
<i>Traditional Approaches: Domain Expertise, Site Content Analysis</i>	136
<i>Keyword Research Tools</i>	138
<i>Determining Keyword Value/Potential ROI</i>	166
<i>Leveraging the Long Tail of Keyword Demand</i>	170
<i>Trending, Seasonality, and Seasonal Fluctuations in Keyword Demand</i>	176
<i>Conclusion</i>	179
6 DEVELOPING AN SEO-FRIENDLY WEBSITE	181
<i>Making Your Site Accessible to Search Engines</i>	181
<i>Creating an Optimal Information Architecture</i>	187
<i>Root Domains, Subdomains, and Microsites</i>	201
<i>Optimization of Domain Names/URLs</i>	207
<i>Keyword Targeting</i>	211
<i>Content Optimization</i>	222
<i>Duplicate Content Issues</i>	226
<i>Controlling Content with Cookies and Session IDs</i>	234
<i>Content Delivery and Search Spider Control</i>	238
<i>Redirects</i>	254
<i>Content Management System (CMS) Issues</i>	263
<i>Optimizing Flash</i>	268
<i>Best Practices for Multilanguage/Country Targeting</i>	275
<i>Conclusion</i>	278
7 CREATING LINK-WORTHY CONTENT AND LINK MARKETING	279
<i>How Links Influence Search Engine Rankings</i>	279
<i>Further Refining How Search Engines Judge Links</i>	289
<i>The Psychology of Linking</i>	295
<i>Types of Link Building</i>	296
<i>Choosing the Right Link-Building Strategy</i>	309
<i>More Approaches to Content-Based Link Acquisition</i>	314
<i>Incentive-Based Link Marketing</i>	319
<i>How Search Engines Fight Link Spam</i>	319
<i>Social Networking for Links</i>	321
<i>Conclusion</i>	330
8 OPTIMIZING FOR VERTICAL SEARCH	333
<i>The Opportunities in Vertical Search</i>	333

<i>Optimizing for Local Search</i>	338
<i>Optimizing for Image Search</i>	346
<i>Optimizing for Product Search</i>	352
<i>Optimizing for News, Blog, and Feed Search</i>	355
<i>Others: Mobile, Video/Multimedia Search</i>	366
<i>Conclusion</i>	372
9 TRACKING RESULTS AND MEASURING SUCCESS	373
<i>Why Measuring Success Is Essential to the SEO Process</i>	374
<i>Measuring Search Traffic</i>	376
<i>Tying SEO to Conversion and ROI</i>	388
<i>Competitive and Diagnostic Search Metrics</i>	397
<i>Key Performance Indicators for Long Tail SEO</i>	435
<i>Conclusion</i>	437
10 DOMAIN CHANGES, POST-SEO REDESIGNS, AND TROUBLESHOOTING	439
<i>The Basics of Moving Content</i>	439
<i>Maintaining Search Engine Visibility During and After a Site Redesign</i>	444
<i>Maintaining Search Engine Visibility During and After Domain Name Changes</i>	445
<i>Changing Servers</i>	447
<i>Hidden Content</i>	448
<i>Spam Filtering and Penalties</i>	455
<i>Content Theft</i>	467
<i>Changing SEO Vendors or Staff Members</i>	470
<i>Conclusion</i>	472
11 HONING THE CRAFT: SEO RESEARCH AND STUDY	473
<i>SEO Research and Analysis</i>	473
<i>Competitive Analysis</i>	482
<i>Using Search-Engine-Supplied SEO Tools</i>	487
<i>The SEO Industry on the Web</i>	500
<i>Participation in Conferences and Organizations</i>	504
<i>Conclusion</i>	506
12 BUILD AN IN-HOUSE SEO TEAM, OUTSOURCE IT, OR BOTH?	507
<i>The Dynamics and Challenges of Using In-House Talent Versus Outsourcing</i>	507
<i>Solutions for Small Organizations</i>	509
<i>Working with Limited Resources/Budget</i>	511
<i>Solutions for Large Organizations</i>	514
<i>Hiring SEO Talent</i>	517
<i>The Case for Working with an Outside Expert</i>	519
<i>Selecting an SEO Firm/Consultant</i>	521
<i>Mixing Outsourced SEO with In-House SEO Teams</i>	528
<i>Building a Culture of SEO into Your Organization</i>	529
<i>Conclusion</i>	530
13 AN EVOLVING ART FORM: THE FUTURE OF SEO	531
<i>The Ongoing Evolution of Search</i>	532
<i>More Searchable Content and Content Types</i>	538
<i>Search Becoming More Personalized and User-Influenced</i>	541

<i>Increasing Importance of Local, Mobile, and Voice Recognition Search</i>	546
<i>Increased Market Saturation and Competition</i>	552
<i>SEO As an Enduring Art Form</i>	554
<i>Conclusion</i>	555
INDEX	557

FOREWORD

MORE THAN A DECADE HAS PASSED SINCE THE WORLD WIDE WEB HAS RISEN TO prominence in nearly all aspects of our lives, but as with nearly all significant technology-driven shifts in our society, most businesses have been slow to react. If you've already put your business online and begun the journey that is your ongoing, online conversation with customers, congratulations!

But if you count yourself as one of those still in the slow lane, fret not. There's never a bad time to get started, and with this book in hand, you're already well on your way.

In fact, starting now might even be to your benefit—over the past decade or so, much has been learned, and many mistakes have been made. New technologies have risen to prominence (Facebook and Twitter come to mind), and old ones have fallen by the wayside. The web is far more mature, and the rules of the road are a bit clearer. In addition, an entire industry of professionals in search optimization and marketing has also matured, and stands ready to help.

Just five years ago a hotshot startup with a funny name went public and, armed with a customer base in the hundreds of thousands and a user base in the tens of millions, proceeded to grow faster than any company in history. In less than a generation, Google has become a cultural phenomenon, a lightning rod for controversy, and a fundamental part of any intelligent businessperson's customer strategy.

But Google is really a proxy for something larger—a new, technologically mediated economy of conversation between those who are looking for products, services, and information, and

those who might provide them. The vast majority of our customers, partners, and colleagues are increasingly fluent in this conversation, and use it to navigate their personal and professional lives.

The shorthand for this interaction is “search,” and like most things worth understanding, learning to leverage search takes practice. In fact, it’s more accurate to put it this way: learning to leverage search is a practice—an ongoing, iterative practice, a process that, once begun, never really finishes. Your customers are out there, asking Google and other search engines questions that by all rights should lead them to your digital doorstep. The question is, are you ready to welcome them?

Think of search as another way to have a conversation with a good customer or prospective customer. The skills you naturally have—at describing your business and its merits relative to competitors, your approach to service, the ecosystem in which your business lives—are skills you should translate to the practice of SEO. It can be daunting and frustrating, but then again, so is starting and running a business. Those who are willing to do the extra work will prosper. Those who stay on the sidelines risk failure.

The days of putting an ad in the Yellow Pages and waiting by the phone are over. With search, everyone’s number is listed—if they have a web site, that is. But not everyone will show up as an answer to a customer query. Learning how to make sure your business shines is no longer an option—it’s table stakes in a game you’ve already decided to play, simply by hanging out a shingle.

So why not play to win? Even if you decide you don’t want to go it alone—and who could blame you?—and you hire an expert to guide you through, understanding the art of SEO will make you a better client for whomever you hire. Speaking from experience, there’s nothing better than working with someone who understands the basics of your practice.

Make no mistake, at the end of the day SEO is an art, one informed by science, experience, and a healthy dose of trial and error. The sooner you get started, the better you and your business will become. The book in your hands is a meticulous volume written by some of the brightest minds and most experienced hands in the SEO industry. Read on, and enjoy the journey!

—John Battelle
Ross, California, September 2009

P R E F A C E

THE BOOK BEFORE YOU IS DESIGNED TO BE A COMPLETE AND THOROUGH EDUCATION on search engine optimization for SEO practitioners at all levels. Think of it as SEO 101, SEO 102, and SEO 103.

We were inspired to create it because we have not previously seen a comprehensive work on the topic of SEO, and we believe that it is very much needed in our industry. Our goal has been to help simplify a very complex, layered topic and to make it easier for people to grasp, as well as to make it easier to focus on the most important aspects of SEO for individual businesses.

As a group we have over 30 years experience working on SEO projects. This means that we have seen how SEO works over a relatively long period of time, across thousands of different web sites. Any one of us could have written this book individually (in fact, one of us tried to), but we discovered that by working together we have been able to create something of much greater value for you, the SEO artist.

Who Should Read This Book

People who are involved in SEO at any level should consider this book invaluable. This includes web developers, development managers, marketing people, and key business personnel. If SEO is not your profession, then this book may serve primarily as a reference. However, if you are or want to become an SEO practitioner, you will likely want to read it from cover to cover.

An experienced SEO veteran will find this volume useful as an extensive reference to support ongoing SEO engagements, both internally, within an in-house SEO group or SEO consultancy, and externally, with SEO clients. Finally, the book will serve as a refresher course for working SEO practitioners, from the novice to the professional.

How to Use This Book

After reading the entire text, a new SEO practitioner will have been exposed to all aspects of the art of SEO and will have laid the necessary groundwork for beginning to develop his SEO expertise.

Conventions Used in This Book

The following typographical conventions are used in this book:

Italic

Indicates new terms, URLs, email addresses, filenames, and file extensions.

Constant width

Used for program listings, as well as within paragraphs to refer to program elements such as variable or function names, databases, data types, environment variables, statements, and keywords.

Constant width italic

Shows text that should be replaced with user-supplied values or by values determined by context.

Using Code Examples

This book is here to help you get your job done. In general, you may use the code in this book in your programs and documentation. You do not need to contact us for permission unless you're reproducing a significant portion of the code. For example, writing a program that uses several chunks of code from this book does not require permission. Selling or distributing a CD-ROM of examples from O'Reilly books does require permission. Answering a question by citing this book and quoting example code does not require permission. Incorporating a significant amount of example code from this book into your product's documentation does require permission.

We appreciate, but do not require, attribution. An attribution usually includes the title, author, publisher, and ISBN. For example: "*The Art of SEO* by Eric Enge, Stephan Spencer, Rand Fishkin, and Jessie C. Stricchiola. Copyright 2010 Eric Enge, Stephan Spencer, Rand Fishkin, and Jessie Stricchiola, 978-0-596-51886-8."

If you feel your use of code examples falls outside fair use or the permission given above, feel free to contact us at permissions@oreilly.com.

Safari® Books Online



Safari Books Online is an on-demand digital library that lets you easily search over 7,500 technology and creative reference books and videos to find the answers you need quickly.

With a subscription, you can read any page and watch any video from our library online. Read books on your cell phone and mobile devices. Access new titles before they are available for print, and get exclusive access to manuscripts in development and post feedback for the authors. Copy and paste code samples, organize your favorites, download chapters, bookmark key sections, create notes, print out pages, and benefit from tons of other time-saving features.

O'Reilly Media has uploaded this book to the Safari Books Online service. To have full digital access to this book and others on similar topics from O'Reilly and other publishers, sign up for free at <http://my.safaribooksonline.com>.

How to Contact Us

Please address comments and questions concerning this book to the publisher:

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
800-998-9938 (in the United States or Canada)
707-829-0515 (international or local)
707-829-0104 (fax)

We have a web page for this book, where we list errata, examples, and any additional information. You can access this page at:

<http://www.oreilly.com/catalog/9780596518868>

The authors also have a companion website at:

<http://www.artofseobook.com>

To comment or ask technical questions about this book, send email to:

bookquestions@oreilly.com

For more information about our books, conferences, Resource Centers, and the O'Reilly Network, see our website at:

<http://www.oreilly.com>

Acknowledgments

We would like to thank comScore, Hitwise, and Nielsen Online for contributing volumes of data to the book. Others who contributed their time or support to our efforts include:

Hamlet Batista—review

John Biundo—review

Jessica Bowman—in-house

Jim Boykin—tools/training access

Matt Cutts—review

Geraldine DeRuiter—outline development and editing

Jon Henshaw—tools access

Brian Klais—metrics

Jill Kocher—review

Chris Smith—local/mobile search

SEOMoz staff —PRO guides

Jeremy Schoemaker—tools access

Danny Sullivan—for his role in launching this industry

Dana Todd—witty wisdom

Aaron Wall—tools/training access

Richard Zwicky—metrics and tools access

CHAPTER ONE

The Search Engines: Reflecting Consciousness and Connecting Commerce

SEARCH HAS BECOME INTEGRATED INTO THE FABRIC OF OUR SOCIETY. With more than 12 billion searches being performed each month as of January 2009 (according to comScore), approximately 400 million web searches are performed every day. This means that on average more than 4,500 searches are performed every single second of every day.

As Google owns approximately 65% of the search market share, Google's search technology handles more than 2,900 searches per second. In addition, users have grown to expect that the responses to their search queries will be returned in less than one second.

Now people can obtain information in mere seconds—information that 20 years ago would have required a trip to the library, a search through the Dewey Decimal System, and a foot search through halls of printed volumes, a process that could easily have consumed two hours or more. Through the new channel of search, people can also conduct much of their shopping, banking, and social transactions online—something that has changed the way our global population lives and interacts.

This dramatic shift in behavior represents what investors like to label a *disruptive event*—an event that has changed something in a fundamental way. Search engines are at the center of this disruptive event, and having a business's website rank well in the search engines when people are looking for the service, product, or resource it provides is critical to the survival of that business. However, like most paths to success, obtaining such prime search result real

estate is not a simple matter, but it is one that this book aims to deconstruct and demystify as we examine, explain, and explore the ever-changing art of search engine optimization (SEO).

The Mission of Search Engines

Search engines generate revenue primarily through paid advertising. The great majority of this revenue comes from a pay-per-click (or cost-per-click) model, in which the advertisers pay only for users who click on their ads.

Since web searchers are free to use any of the many available search engines on the Web to find what they are seeking, the burden is on the search engines to develop a relevant, fast, and fresh search experience. For the most part, search engines accomplish this by being *perceived* as having the most relevant results and delivering them the fastest, as users will go to the search engine they think will get them the answers they want in the least amount of time.

As a result, search engines invest a tremendous amount of time, energy, and capital in improving their relevance. This includes performing extensive studies of user responses to their search results, comparing their results against those of other search engines, conducting eye-tracking studies (discussed later in this chapter), and conducting PR and marketing campaigns.

Because the search engines' success depends so greatly on the relevance of their search results, manipulations of search engine rankings that result in non-relevant results (generally referred to as *spam*) are dealt with very seriously. Each major search engine employs a team of people who focus solely on finding and eliminating spam from their search results. This matters to SEO practitioners because they need to be careful that the tactics they employ will not be seen as spam by the search engines and carry the risk of resulting in penalties for the websites they work on.

The Market Share of Search Engines

Figure 1-1 shows the U.S. market share for search engines in January 2009, according to comScore. As you can see, Google is the dominant search engine on the Web in the United States.

In many European countries, the disparity is even greater. However, in some markets Google is not dominant. In China, for instance, Baidu is the leading search engine. The result is that in most world markets, a heavy focus on SEO is a smart strategy for Google.

The Human Goals of Searching

The basic goal of a human searcher is to obtain information relevant to her inquiry. However, searcher inquiries can take many different forms. One of the most important elements to building an online marketing strategy for a website around SEO and search rankings is

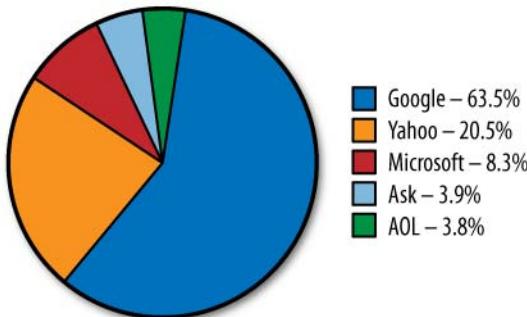


FIGURE 1-1. Search engine market share

developing a thorough understanding of the psychology of your target audience. Once you understand how the average searcher, and more specifically, your target market, uses search engines, you can more effectively reach and keep those users.

Search engine usage has evolved over the years, but the primary principles of conducting a search remain largely unchanged. The following steps comprise most search processes:

1. Experience the need for an answer, solution, or piece of information. For example, the user may be looking for a website (navigational query) to buy something (transactional query) or to learn something (informational query). We will discuss this in more detail in the following section.
2. Formulate that need in a string of words and phrases (the query). Most people formulate their queries in one to three words. ComScore data from March 2009 shows an average query length of 2.9 words. A more detailed look shows the following percentages of searches per word length (see [Table 1-1](#)).

TABLE 1-1. Searches by query length (comScore)

Words	Percent of searches
1	25.32%
2	24.96%
3	19.80%
4	13.17%
5	7.53%
6	4.04%
7	2.15%
8	1.19%

Data from Hitwise for this book shows a similar distribution of search query lengths (see [Table 1-2](#)).

TABLE 1-2. Searches by query length (Hitwise)

Percentage of U.S. clicks by number of keywords				
Subject	February 2008	January 2009	February 2009	Year-over-year % change
One word	21.04%	20.29%	20.48%	-3%
Two words	24.73%	23.65%	23.47%	-5%
Three words	21.84%	21.92%	21.68%	-1%
Four words	14.53%	14.89%	14.98%	3%
Five words	8.29%	8.68%	8.72%	5%
Six words	4.38%	4.65%	4.71%	8%
Seven words	2.29%	2.49%	2.51%	10%
Eight or more words	2.90%	3.43%	3.47%	20%

Note: data is based on four-week rolling periods (ending February 28, 2009; January 31, 2009; and March 1, 2008) from the Hitwise sample of 10 million U.S. Internet users.

Source: Hitwise, an Experian company

3. Execute the query, check the results, see whether you got what you wanted, and if not, try a refined query.

When this process results in the satisfactory completion of a task, a positive experience is created for the user, the search engine, and the site providing the information or result.

Who Searches and What Do They Search For?

ComScore reported that the number of search queries performed on the Web was approximately 12.6 billion across all engines in December 2008.

ComScore data shows that just under 79 million people in the United States were using a search engine on a given day in January 2009. Search engine users were slightly more likely to be women than men (50.4% versus 49.6%). According to comScore, Internet usage increases with household income, as per the data shown in [Table 1-3](#) for January 2009.

TABLE 1-3. Internet users by household income

Household income	Internet users
\$15,000–\$24,999	5,792
\$25,000–\$39,999	16,108

Household income	Internet users
\$40,000–\$59,999	39,716
\$60,000–\$74,999	20,947
\$75,000–\$99,000	28,995
\$100,000 or more	44,627

You can find additional data from studies, surveys, and white papers on Search Engine Land's [Stats & Behaviors page](#).

All of this research data leads us to some important conclusions about web search and marketing through search engines. For example:

- Search is very, very popular. It reaches nearly every online American and billions of people around the world.
- Google is the dominant player in most world markets.
- Users tend to use short search phrases, but these are gradually getting longer.
- Search covers all types of markets.

Search is undoubtedly one of the best and most important ways to reach consumers and build a business, no matter the size, reach, or focus.

Determining Searcher Intent: A Challenge for Both Marketers and Search Engines

Good marketers are empathetic. Smart SEO practitioners and the search engines have a common goal of providing searchers with results that are relevant to their queries. Therefore, a crucial element to building an online marketing strategy around SEO and search rankings is to understand your audience. Once you grasp how your target market searches for your service, product, or resource, you can more effectively reach and keep those users.

Search engine marketers need to be aware that search engines are *tools*—resources driven by intent. Using the search box is fundamentally different from entering a URL into the address bar, clicking on a bookmark, or picking a link on your start page to go to a website; it is unique from a click on the “stumble” button in your StumbleUpon toolbar or a visit to your favorite blog. Searches are performed with *intent*; the user wants to find something in particular, rather than just land on it by happenstance.

What follows is an examination of the different types of queries, their categories, characteristics, and processes.

Navigational Queries

Navigational searches are performed with the intent of surfing directly to a specific website. In some cases, the user may not know the exact URL, and the search engine serves as the “White Pages.” [Figure 1-2](#) shows an example of a navigational query.



FIGURE 1-2. Navigational query

Opportunities: Pull searcher away from destination; get ancillary or investigatory traffic.

Average value: Generally low, with the exception of navigational searches on the publisher’s own brand, where the value is very high as these types of searches tend to lead to very high conversion rates.

Informational Queries

Informational searches involve a huge range of queries—for example, local weather, maps and directions, details on the latest Hollywood awards ceremony, or just checking how long that trip to Mars really takes. Informational searches are primarily non-transaction-oriented (although they can include researching information about a product or service); the information itself is the goal and no interaction beyond clicking and reading is required. [Figure 1-3](#) shows an example of an informational query.

Opportunities: Brand searchers with positive impressions of your site, information, company, and so on; attract inbound links; receive attention from journalists/researchers; potentially convert to sign up or purchase.

Average value: Middling. Note, though, that informational queries that are focused on researching commercial products or services can have high value.

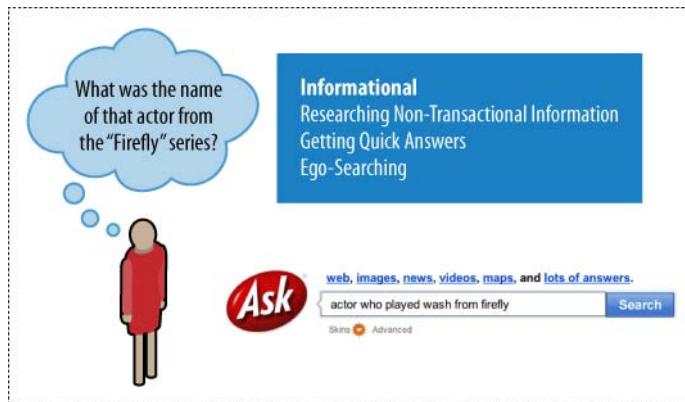


FIGURE 1-3. Informational query

Transactional Queries

Transactional searches don't necessarily involve a credit card or wire transfer. Signing up for a free trial account at CookIllustrated.com, creating a Gmail account, paying a parking ticket, or finding the best local Mexican cuisine for dinner tonight are all transactional queries.

[Figure 1-4](#) shows an example of a transactional query.



FIGURE 1-4. Transactional query

Opportunities: Achieve transaction (financial or other).

Average value: Very high.

Research by Pennsylvania State University and the Queensland University of Technology (http://ist.psu.edu/faculty_pages/jjansen/academic/pubs/jansen_user_intent.pdf) shows that more than

80% of searches are informational in nature and only about 10% of searches are navigational or transactional.

The researchers went further and developed an algorithm to automatically classify searches by query type. When they tested the algorithm, they found that it was able to correctly classify queries 74% of the time. The difficulty in classifying the remaining queries was vague user intent, that is, the query could have multiple meanings. Here are some URLs that point to additional academic research on this topic:

- <http://www.sigir.org/forum/F2002/broder.pdf>
- <http://www.strategynode.com/how-to-determine-if-a-search-query-is-navigational-informational-or-transactional>

When you are building keyword research charts for clients or on your own sites, it can be incredibly valuable to determine the intent of each of your primary keywords. **Table 1-4** shows some examples.

TABLE 1-4. Sample search queries and intent

Term	Queries	Intent	\$\$ value
<i>Beijing Airport</i>	980	Nav	Low
<i>Hotels in Xi'an</i>	2,644	Info	Mid
<i>7-Day China tour package</i>	127	Trans	High
<i>Sichuan jellyfish recipe</i>	53	Info	Low

This type of analysis can help to determine where to place ads and where to concentrate content and links.

Hopefully, this data can help you to think carefully about how to serve different kinds of searchers, based on their individual intents, and then concentrate your efforts in the best possible areas.

Although informational queries are less likely to immediately convert into sales, this does not necessarily mean you should forego rankings on such queries. If you are able to build a relationship with users who find your site after an informational query, they may be more likely to come to you to make the related purchase at a later date.

One problem is that when most searchers frame their search query they provide very limited data to the search engine, usually in just one to three words. Since most people don't have a keen understanding of how search engines work, they can often provide a query that is too general or in a way that does not provide the search engine (or the marketer) with what it needs to determine their intent.

For this reason, general queries are important to most businesses because they often get the brand and site on the searcher's radar, and this initiates the process of building trust with the user. Over time, the user will move on to more specific searches that are more transactional or navigational in nature.

If, for instance, companies buying pay-per-click (PPC) search ads bought only the high-converting navigational and transactional terms and left the informational ones to competitors, they would lose market share to those competitors. During several days, a searcher may start with *digital cameras* and then hone in on *canon g10*, and buy from the store that showed up for digital cameras and pointed her in the direction of the Canon G10 model.

Given the general nature of how query sessions start, though, determining intent is quite difficult, and can result in searches being performed where the user does not find what she wants, even after multiple tries. An [August 2007 Foresee/ACSI Report for eMarketer](#) found that 75% of search engine and portal users were satisfied with their experiences. In a breakdown by property, 79% of Yahoo! users, 78% of Google users, and 75% of both Live Search (Microsoft's web search property, which has since been renamed to Bing) and [Ask.com](#) users reported being satisfied.

Based on this later study, more than 20% of users did not find what they were looking for. This suggests that there is plenty of room to improve the overall search experience. As an SEO practitioner, you should be aware that many of the visitors that you succeed in attracting to your site may have arrived for the wrong reasons (i.e., they were really looking for something else), and these visitors are not likely to help your business goals.

How People Search

Search engines invest significant resources into understanding how people use search, enabling them to produce better (i.e., faster, fresher, and more relevant) search engine results. For website publishers, the information regarding how people use search can be used to help improve the usability of the site as well as search engine compatibility.

Data from comScore provides some great insight into what people actually search for when they perform a search. [Table 1-5](#) shows a breakdown of many of the major categories that people's Internet searches fall into, based on comScore data for October 2008.

TABLE 1-5. Searches by market segment

Parent category name	Percent of total searches
Directories/Resources	16.60%
Retail	11.86%
Entertainment	11.54%
Services	6.63%
Education	4.59%
Conversational Media	4.04%
Government	3.87%
Health	3.38%
Games	3.26%
News/Information	3.06%
Hobbies/Lifestyle	3.05%
Business/Finance	2.94%
Travel	2.21%
Community	1.94%
Regional/Local	1.87%
Sports	1.78%
Technology	1.73%
Automotive	1.67%
Real Estate	1.43%
Career Services and Development	1.12%
Telecommunications	0.78%
Auctions	0.57%
Portals	0.56%
ISP	0.38%
Gambling	0.27%
Business to Business	0.25%

This shows that people search across a very wide number of categories. Search engines are used to find information in nearly every portion of our lives. In addition, user interactions with

search engines can be a multistep process. Witness the user search session documented by Microsoft and shown in [Figure 1-5](#).

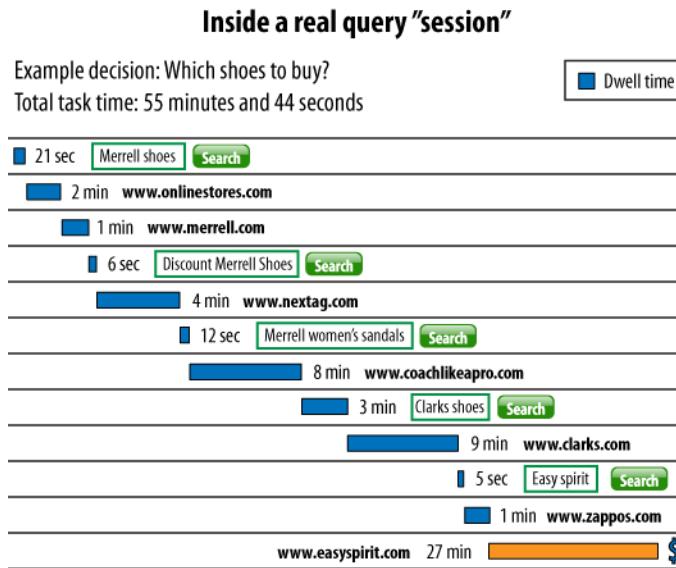


FIGURE 1-5. Merrell shoes user search session

In this sequence, the user performs five searches over a 55+ minute period before making a final selection. The user is clearly trying to solve a problem and works at it in a persistent fashion until the task is done.

However, it is increasingly common that search sessions of this type can take place over days. A 2007 study of e-commerce sites by ScanAlert showed that 30% of online transactions occurred more than 24 hours after the initial search (<http://searchenginewatch.com/3626363>).

This means people are thinking about their tasks in stages. As in our Merrell shoes example in [Figure 1-5](#), people frequently begin with a general term and gradually get more specific as they get closer to their goal. They may also try different flavors of general terms. In [Figure 1-5](#), it looks like the user did not find what she wanted when she searched on *Merrell shoes*, so she then tried *discount Merrell shoes*. You can then see her refine her search in the process, until she finally settles on Easy Spirit as the type of shoe she wants.

This is just one example of a search sequence, and the variety is endless. [Figure 1-6](#) shows another search session, once again provided courtesy of Microsoft.

In this search session, the user has a health concern. This particular user starts with a five-word search, which suggests that she may have some experience using search engines. At 3:01 her search on *headache pregnant 3rd trimester* leads her to Answers.yahoo.com. After visiting this site, her search suddenly gets more specific.

Typical session: health

10 unique queries
7 partial re-queries and refinements
57 minutes

2:58	Google	nausea headache pregnant what's wrong
		pregnancy.families.com
2:58		nausea headache pregnant 3rd trimester
	2 min	answers.yahoo.com/question
3:00		Define malaise
	1 min	www.chefjournal.org/content
3:01		Nausea headache pregnant 3rd trimester Headache pregnant 3rd trimester
	5 min	answers.yahoo.com/question
3:06		gestational diabetes gestational diabetes personal story
	16 min	www.experienceproject.com
3:22		pregnancy complications stories
	4 min	www.diabetictmommy.com/articles.html
3:26		gestational diabetes itchy skin
	14 min	www.steadyhealth.com/Diabetes
3:40		first signs of gestational diabetes
	15 min	www.articlebase.com; www.man-health-fitness-solutions.com; www.healthspy.net; www.diabeteslife.com/diabetes; www.optimundiabetics.com; www.everydayhealth.com; www.squidoo.com; mascan-files-anti.com/200092/scan

FIGURE 1-6. Health user search session

She begins to focus on gestational diabetes, perhaps because something she saw on [Answers.yahoo.com](#) led her to believe she may have it. The session culminates in a search for *first signs of gestational diabetes*, which suggests that she has concluded that this is quite possibly the issue she is facing.

The session stops there. It may be that at this point the user feels she has learned what she can. Perhaps her next step is to go to her doctor with her concerns, prepared to ask a number of questions based on what she learned.

Our next search session example begins with a navigational search, where the user simply wants to find the site for the travel website Orbitz (see [Figure 1-7](#)). The user's stay there is quite short, and she progresses to a search on *Cancun all inclusive vacation packages*. Following that she searches on a few specific resorts but finally settles on *cancun riviera maya hotels*, after which it appears she may have booked her hotel—the final site visited on that search is [Bookings.occidentalhotels.com](#), and the direction of her searches changes after that.

Typical session: travel

11 unique queries
5 partial re-queries and refinements
33 minutes

19:14 Google	Orbitz travel	1 min www.orbitz.com
19:15 Google	Cancun all inclusive vacation packages	14 min cancun.bookit.com; www.travelocity.com; from bookit.com
19:29 Google	Allegro Playacar - All inclusive	2 min from bookit.com; www.ocidentalhotels.com
19:31 Google	Occidental Grand Xcaret - All inclusive resort	3 min from bookit.com; search.bookit.com; www.ocidentalhotels.com; from bookit.com
19:34 Google	NH Riviera Cancun - All inclusive	1 min www.cancun.com/Hotels/NH-Riviera-Cancun; www.tripadvisor.com; www.travel-center.com
19:35 Google	cancun riviera maya hotels	7 min www.nh-hotels.com; www.cancunandrivieramaya.com; www.nh-hotels.com; www.ocidentalhotels.com; bookings.ocidentalhotels.com
19:42 YAHOO!	Cancun theme park	1 min www.cancun.com/Ecological_Parks/Xcaret
19:43 Live	xcaret	1 min search.live.com/results; search.live.com/video
19:44 Live	xcaret	2 min www.xcaret.com; www.xcaret.com/Gallery.aspx; bookings.ocidentalhotels.com/travel/tinerary.cfm; www.xcaret.com/Tickets/Xcaret_Basic_Entrance.html
19:46 Live	xcaret video	1 min search.live.com/video; www.orbitz.com/App/View; www.orbitz.com/App/ViewPackageDetails

FIGURE 1-7. Travel user search session

At that point, the user begins to look for things to do while she is in Cancun. She conducts a search for *cancun theme park* and then begins to look for information on *xcaret*, a well-known eco park in the area.

Users traverse through countless different scenarios when they are searching for something. Search engines do a lot of modeling of these scenarios to enable them to provide better results to users. The SEO practitioner can benefit from a basic understanding of searcher behavior as well. We will discuss this in more detail in [Chapter 2](#).

How Search Engines Drive Commerce on the Web

People make use of search engines for a wide variety of purposes, with some of the most popular being to research, locate, and buy products. Even with the recession that (officially) hit the U.S. economy in late 2007, and which continues through the publication date of this book, e-commerce sales reported by the U.S. Census Bureau were a healthy \$31.9 billion (<http://www.census.gov/mrts/www/data/pdf/08Q4.pdf>) through 2008.

It is important to note that search and offline behavior have a heavy degree of interaction, with search playing a growing role in driving offline sales. A Yahoo! study from 2007 (<http://searchengineland.com/yahoos-robo-study-search-has-big-impact-on-offline-purchases-11832>) showed the following:

- Online advertising drives \$6 offline (in stores) for every \$1 spent online.
- Search marketing has a greater impact on in-store sales lift than display advertising—three times greater, in fact.

There is also a significant amount of interaction between search and local offline commerce. WebVisible and Nielsen produced a 2007 report on local search (<http://searchengineland.com/survey-search-now-top-resource-for-local-information-12396>) that noted:

- 74% of respondents used search engines to find local business information versus 65% who turned to print Yellow Pages, 50% who used Internet Yellow Pages, and 44% who used traditional newspapers.
- 86% surveyed said they have used the Internet to find a local business, a rise from the 70% figure reported the year before.
- 80% reported researching a product or service online, then making that purchase offline from a local business.

Local search is an increasingly important component of SEO, and one which we will explore in detail in [Chapter 2](#).

Eye Tracking: How Users Scan Results Pages

Research firms [Enquiro](#), [Eyetools](#), and [Didit](#) conducted heat-map testing with search engine users (<http://www.enquiro.com/research/eyetrackingreport.asp>) that produced fascinating results about what users see and focus on when engaged in search activity. [Figure 1-8](#) depicts a heat map showing a test performed on Google. The graphic indicates that users spent the most amount of time focusing their eyes in the top-left area where shading is the darkest.

Published in November 2006, this particular study perfectly illustrates how little attention is paid to results lower on the page versus those higher up, and how users' eyes are drawn to bold keywords, titles, and descriptions in the natural ("organic") results versus the paid search listings, which receive comparatively little attention.

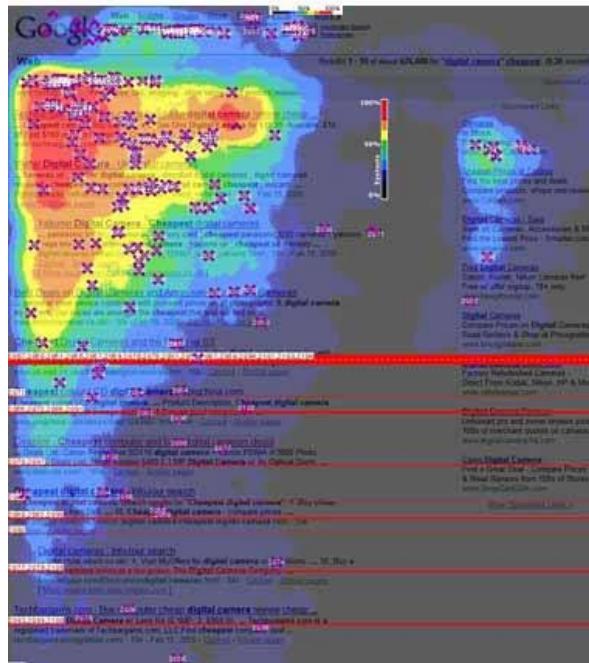


FIGURE 1-8. Enquiero eye-tracking results

This research study also showed that different physical positioning of on-screen search results resulted in different user eye-tracking patterns. When viewing a standard Google results page, users tended to create an “F-shaped” pattern with their eye movements, focusing first and longest on the upper-lefthand corner of the screen; moving down vertically through the first two or three results; moving across the page to the first paid page result; moving down another few vertical results; and then moving across again to the second paid result. (This study was done only on left-to-right language search results—Chinese, Hebrew, etc. results would be different.)

In May 2008, Google introduced the notion of Universal Search. This was a move from simply showing the 10 most relevant web pages (now referred to as “10 blue links”) to showing other types of media, such as videos, images, news results, and so on, as part of the results in the base search engine. The other search engines followed suit within a few months, and the industry now refers to this general concept as *Blended Search*.

Blended Search, however, creates more of a chunking effect, where the chunks are around the various rich media objects, such as images or video. Understandably, users focus on the image first. Then they look at the text beside it to see whether it corresponds to the image or video thumbnail (which is shown initially as an image). Based on an updated study published

by Enquiro in September 2007, [Figure 1-9](#) shows what the eye-tracking pattern on a Blended Search page looks like.



FIGURE 1-9. Enquiro eye-tracking results, Blended Search

Users' eyes then tend to move in shorter paths to the side, with the image rather than the upper-left-corner text as their anchor. Note, however, that this is the case only when the image is placed above the fold, so that the user can see it without having to scroll down on the page. Images below the fold do not influence initial search behavior until the searcher scrolls down.

This study is a vivid reminder of how important search engine results pages (SERPs) really are. And as the eye-tracking research demonstrates, "rich" or "personalized" search, as it evolves, will alter users' search patterns even more: there will be more items on the page for them to focus on, and more ways for them to remember and access the search listings. Search marketers need to be prepared for this as well.

Click Tracking: How Users Click on Results, Natural Versus Paid

By now, you should be convinced that you want to be on the top of the SERPs. It never hurts to be #1 in the natural search results.

In a bit of contrast to that, data shows that you may not want to be #1 in the paid search results, because the resulting cost to gain the #1 position in a PPC campaign can reduce the total net margin on your campaign. A [study released by AdGooroo in June 2008](#) found that:

Bidding for top positions usually makes financial sense only for high-budget, brand-name advertisers. Most other advertisers will find the optimal position for the majority of their keywords to lie between positions 5–7.

Of course, many advertisers may seek the #1 position in paid search results for a number of reasons. For example, if they have a really solid backend on their website and are able to make money when they are in the #1 position, they may well choose to pursue it. Nonetheless, the data from the survey suggests that there are many organizations for which being #1 in paid search does not make sense.

Even if your natural ranking is #1, you can still increase the ranking page's click rate by having a sponsored ad above it or in the right column. The AdGooroo survey showed that a prominent paid ad on the same search results page makes your #1 natural ranking receive 20% more clicks.

Distribution of Search Results and Traffic

To start breaking this down a bit, [Figure 1-10](#) shows the screen real estate occupied by the two types of search results.

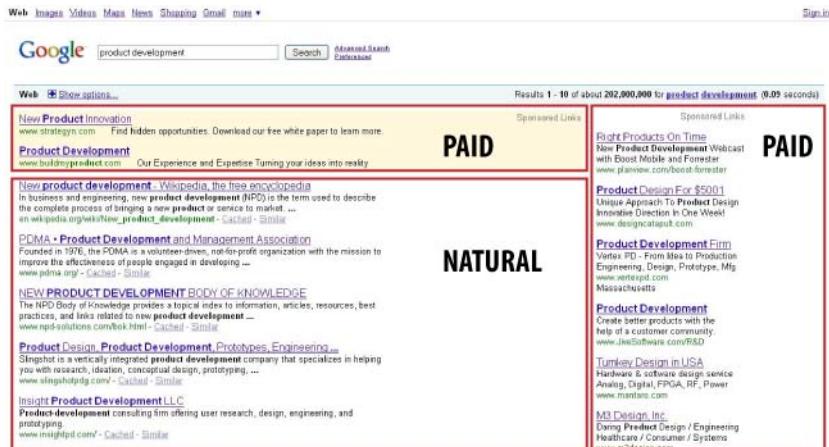


FIGURE 1-10. Paid and natural search results

This example from Google shows how the paid results appear above and to the right of the natural search results. Note that Google often does not show paid results above the natural results, in which case the paid results show up only on the right.

Your position in the results has a huge impact on the traffic you will receive. Here is some data about that:

- An [April 2006 study by iProspect and Jupiter Research](#) found that:
 - 62% of search engine users click on a search result within the first page of results, and 90% within the first three pages.
 - 41% of search engine users who continue their search when not finding what they seek report changing their search term and/or search engine if they do not find what they're looking for on the first page of results; 88% report doing so after three pages.
 - 36% of users agree that “seeing a company listed among the top results on a search engine makes me think that the company is a top one within its field.”

- A study on data leaked from AOL's search query logs in August 2006 (<http://www.webuildpages.com/jim/click-rate-for-top-10-search-results/>) reveals the impact of position on the search results on the percentage of clicks received, as shown in Figure 1-11. In addition, the first 10 results received 89.71% of all click-through traffic; the next 10 results (normally listed on the second page of results) received 4.37%; the third page 2.42%; and the fifth page 1.07%. All other pages of results received less than 1% of total search traffic clicks.

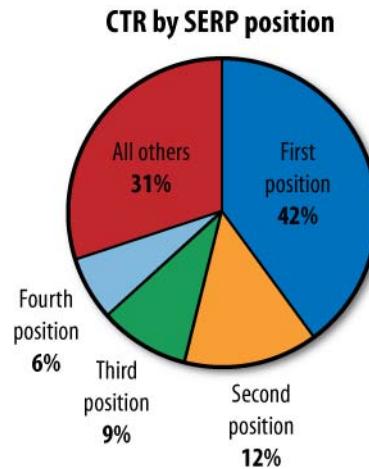


FIGURE 1-11. Click-through rate by SERP position

Why are searchers blind to relevant results farther down the page? Is this due to the “implied endorsement” effect, whereby searchers tend to simply trust the search engine to point them to the right thing? Or is it just the way humans are wired, to make snap decisions, as Malcolm Gladwell insightfully explains in his book *Blink* (Little Brown)?

According to the study, 72% of searchers click on the first link of interest, whereas 25.5% read all listings first and then decide which one to click. Both effects (implied endorsement and rapid cognition) most likely play a role in searcher behavior.

Different Intents and Effects of Listings in Paid Versus Natural Results

The AOL data in Figure 1-11 demonstrated that natural results get the lion’s share of click results. Further data from the Enquiro, Didit, and Eyetools eye-tracking study shows which results users notice when looking at a search results page (see Table 1-6).

TABLE 1-6. Visibility of natural search results

Rank	Visibility
1	100%
2	100%
3	100%
4	85%
5	60%
6	50%
7	50%
8	30%
9	30%
10	20%

Table 1-7 shows the percentage of participants looking at a listing in this location in the paid results.

TABLE 1-7. Visibility of paid search results

Rank	Visibility
1	50%
2	40%
3	30%
4	20%
5	10%
6	10%
7	10%
8	10%

Notice how the visibility of a listing in the natural results is double or more (up to six times) of the visibility of the same position in the paid results. For example, in position 5 only 60% of users ever even notice the natural search result. But the paid search results fare even worse, with only 10% of users noticing the result in the fifth position.

Here are some additional takeaways from the Enquiero et al. study:

- 85% of searchers click on natural results.

- The top four sponsored slots are equivalent in views to being ranked at 7–10 in natural search in terms of visibility and click-through.
- This means if you need to make a business case for natural search, then (assuming you can attain at least the #3 rank in natural search for the same keywords you bid on) natural search could be worth two to three times your PPC results.

It is interesting to note that in spite of this data, companies are much more likely to spend money on PPC than SEO. For example, on June 30, 2008, Jupiter Research released a report (http://news.cnet.com/8301-1023_3-9980927-93.html) showing that search advertising should continue to be the largest category of online ad spending, growing from \$9.1 billion in 2007 to \$20.9 billion in 2013. That is significant growth.

During the same period, [SEMPO data via Massimo Burgio](#) shows that spending on SEO was \$1.3 billion, with 11% of search-related budgets going to SEO and 87% to PPC (with another 1.4% to PPC technologies and <1% to paid inclusion).

Clearly, the PPC model is easier for companies to understand because it is more similar to traditional direct marketing methods than SEO is. The ROI of PPC can be tracked and demonstrated more reliably than SEO; thus, to date it has been considered more accountable as a marketing channel. However, as budgets are tightening and the focus is shifting to the highest ROI search investments, the focus is increasingly on SEO.

Interaction Between Natural and Paid Search

ICrossing published a report (<http://www.icrossing.com/articles/Search%20Synergy%20Report.pdf>) that showed a strong synergy between natural and paid search. The study shows what happens when you incorporate natural search into an existing paid search campaign and compare its performance to the performance of the sole paid search campaign. [Figure 1-12](#) summarizes the improvement in the results.

The marked improvement in click-through rate intuitively makes sense. For years marketers have known that the number of impressions a consumer is exposed to will have a dramatic effect on metrics such as retention and likelihood to buy.

A search page provides you with more than one opportunity to put your name in front of the user. You should take advantage of this if you can. It is also useful to understand the difference between natural and paid search. Although some users do not understand the distinction between natural search results and paid search results, it is a well-accepted belief in the industry that the majority of users recognize paid search results as advertisements.

Other Factors to Consider

There are many other aspects to consider when thinking about search and your business. Here are some interesting examples:

Online performance percentage change when integrating natural search and paid search

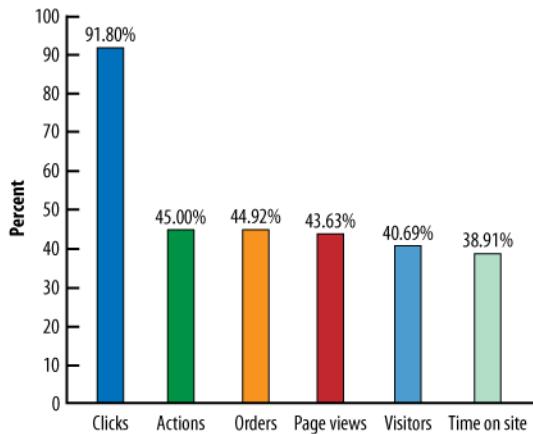


FIGURE 1-12. Interaction between natural and paid search

Number of visits before purchase

A May 2007 study by ScanAlert showed that only 43% of users who made a purchase on a site made that purchase within an hour of their initial visit to the site. So, to translate that a bit, only 43% of users were likely to have made their purchase upon their first visit to the site. In addition, 26% of purchasers made their purchases after three days or more! This suggests that it is very important for publishers to plan on more complex interactions with users across multiple visits to the site.

Presentation changes made by the search engines

In February 2008, Yahoo! announced its [SearchMonkey platform](#). This platform allows publishers and developers to directly influence the search results shown when your web pages are included in those results. [Figure 1-13](#) shows a sample search result that demonstrates this.

Notice how the third listing (Yelp) looks different from the rest of the results. The difference in the look of the results really catches the eye, and can significantly impact click-through rates. And it works, as Amit Kumar of Yahoo! recently stated: “Our tests uncovered that users found these apps useful; in fact, in some cases, we saw a lift in click-through rate of as high as 15 percent.”



FIGURE 1-13. Yahoo!'s SearchMonkey

Conclusion

Search has penetrated the very fabric of global society. The way people work, play, shop, research, and interact has changed forever. Organizations of all kinds (businesses and charities), as well as individuals, need to have a presence on the Web—and they need the search engines to bring them traffic. As our society moves ever closer to a professional consumer (“prosumer”) economy, the ways in which people create, publish, distribute, and ultimately find information and resources on the Internet will continue to be of great importance. This book will investigate further just how search, and therefore search engine optimization, is at the center of the Web and is your key to success in the new web economy.

CHAPTER TWO

Search Engine Basics

IN THIS CHAPTER, WE WILL BEGIN TO EXPLORE HOW SEARCH ENGINES WORK. Building a strong foundation on this topic is essential to understanding the SEO practitioner's craft.

As we discussed in [Chapter 1](#), people have become accustomed to receiving nearly instantaneous answers from search engines after they have submitted a search query. In [Chapter 1](#) we also discussed the volume of queries (more than 4,500 per second), and Google reported in July 2008 that it knew of about 1 trillion pages on the Web (<http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>).

Underlying this enormous data processing task is the complex nature of the task itself. One of the most important things to understand about search engines is that the crawlers (or spiders) used to visit all the web pages across the Web are software programs. Software programs are only as smart as the algorithms used in implementing them, and although artificial intelligence is being increasingly used in those algorithms, web crawling programs still don't have the adaptive intelligence of human beings.

Software programs cannot adequately interpret each of the various types of data that humans can—videos and images, for example, are to a certain extent less readable by a search engine crawler than they are through the eyes of humans. In the section “[Vertical Search Engines](#)” on page 65, we will discuss how advances in image and video search have enabled search engines to inch even closer to human-like understanding.

Understanding Search Engine Results

In the search marketing field, the pages the engines return to fulfill a query are referred to as *search engine results pages* (SERPs). Each engine returns results in a slightly different format and will include *vertical results* (specific content targeted to a query based on certain triggers in the query, which we'll illustrate shortly).

Understanding the Layout of Search Results Pages

Figure 2-1 shows the SERPs in Google for the query *stuffed animals*.

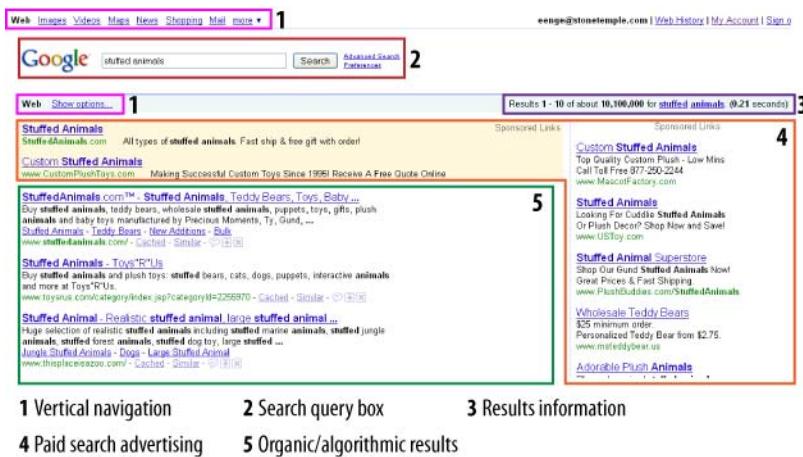


FIGURE 2-1. Layout of Google search results

Figure 2-2 shows Yahoo!'s results for the same query.

Figure 2-3 shows the layout of the results from Microsoft's Bing.

Each unique section represents a snippet of information provided by the engines. Here are the definitions of what each piece is meant to provide:

Vertical navigation

Each engine offers the option to search different verticals, such as images, news, video, or maps. Following these links will result in a query with a more limited index. In

Figure 2-3, for example, you might be able to see news items about stuffed animals or videos featuring stuffed animals.

Search query box

All of the engines show the query you've performed and allow you to edit or reenter a new query from the search results page. Next to the search query box, the engines also

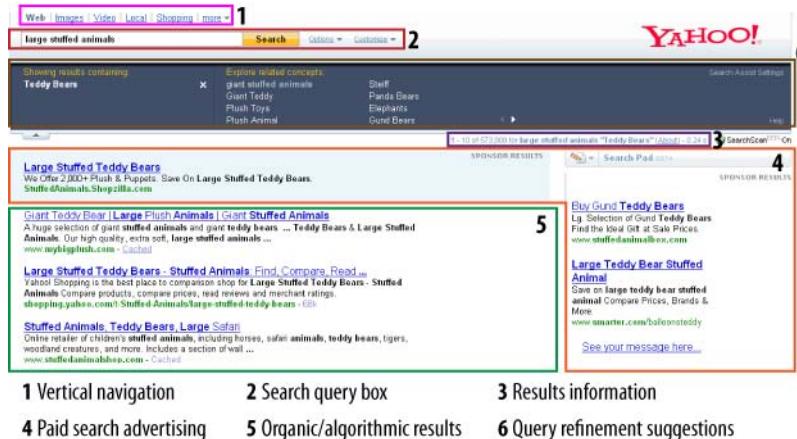


FIGURE 2-2. Layout of Yahoo! search results

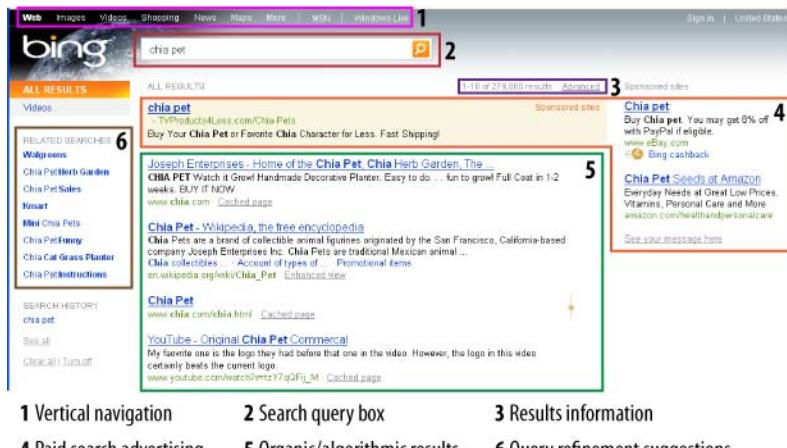


FIGURE 2-3. Layout of Bing search results

offer links to the advanced search page, the features of which we'll discuss later in the book.

Results information

This section provides a small amount of meta information about the results that you're viewing, including an estimate of the number of pages relevant to that particular query (these numbers can be, and frequently are, wildly inaccurate and should be used only as a rough comparative measure).

PPC (a.k.a. paid search) advertising

The “Sponsored Results,” to use Yahoo!’s terms, are text ads purchased by companies that use the various search ad platforms—Google AdWords, Yahoo! Search Marketing, and Microsoft adCenter. The results are ordered by a variety of factors, including relevance (for which click-through rate, use of searched keywords in the ad, and relevance of the landing page are factors in Google) and bid amount (the ads require a maximum bid, which is then compared against other advertisers’ bids).

Natural/organic/algorithmic results

These results are pulled from the search engines’ primary indexes of the Web and ranked in order of relevance and popularity according to their complex algorithms. This area of the results is the primary focus of this section of the book.

Query refinement suggestions

A relatively recent feature, query refinements are now offered by all three engines. The goal of these links is to let users search with a more specific and possibly more relevant query that will satisfy their intent.

In March 2009, Google enhanced the refinements by implementing Orion Technology, based on technology Google acquired in 2006. The goal of this enhancement is to provide a wider array of refinement choices. For example, a search on *principles of physics* will display refinements for the Big Bang, angular momentum, quantum physics, and special relativity.

Be aware that the SERPs are always changing as the engines test new formats and layouts. Thus, the images in Figures 2-1 through 2-3 may be accurate for only a few weeks or months until Google, Yahoo!, and Microsoft shift to new formats.

How Vertical Results Fit into the SERPs

These “standard” results, however, are certainly not all that the engines have to offer. For many types of queries, search engines show *vertical* results, or *instant answers* (which Google collectively refers to as *onebox results*), and include more than just links to other sites to help answer a user’s questions. These types of results present many additional challenges and opportunities for the SEO practitioner.

Figure 2-4 shows an example of these types of results. The query in Figure 2-4 brings back a direct map with an address and the option to get directions. This result is drawn from a Google Maps search and attempts to provide the user with the answer he is seeking directly in the search results.

Figure 2-5 shows another example. The Google search in Figure 2-5 for weather plus a city name returns a direct answer. Once again, the user may not even need to click on a website if all he wanted to know was the temperature.



FIGURE 2-4. Local search result for a business

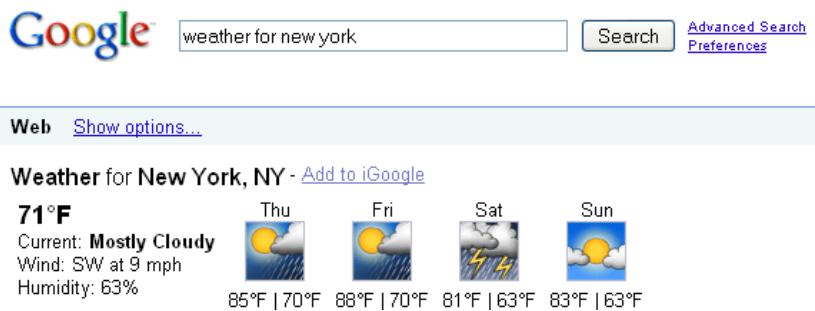


FIGURE 2-5. Weather search on Google

Figure 2-6 is an example of a search for a well-known painter. A Google search for the famous painter Edward Hopper returns image results of some of his most memorable works. This example is a little different from the “instant answers” type of result shown in Figures 2-4 and 2-5. If the user is interested in the first painting shown, he may well click on it to see the painting in a larger size or to get more information about it. For the SEO practitioner, getting placed in this onebox result could be a significant win.

Figure 2-7 shows an example from Yahoo!. A query on Yahoo! for *chicago restaurants* brings back a list of popular dining establishments from Yahoo!'s local portal. High placement in these results has likely been a good thing for Giordanos Pizzeria.

Figure 2-8 is an example of an instant answer result from Yahoo!. Searching Yahoo! to find the number of pounds in a ton brings back an instant answer of 2,000 pounds. Microsoft’s Bing also presents vertical results and instant answer results.



FIGURE 2-6. Google search on an artist's name



FIGURE 2-7. Yahoo! search for Chicago restaurants



FIGURE 2-8. Yahoo! instant answer result

Figure 2-9 is an example of a celebrity search.

The results in Figure 2-9 include a series of images of the famous actor Charlie Chaplin. As a last example, Figure 2-10 is a look at the Bing search results for videos with Megan Fox.

At the top of the search results in Figure 2-10, you get a series of popular videos provided. Click on a video in the results and the video begins playing right there in the search results.



FIGURE 2-9. Bing result for Charlie Chaplin

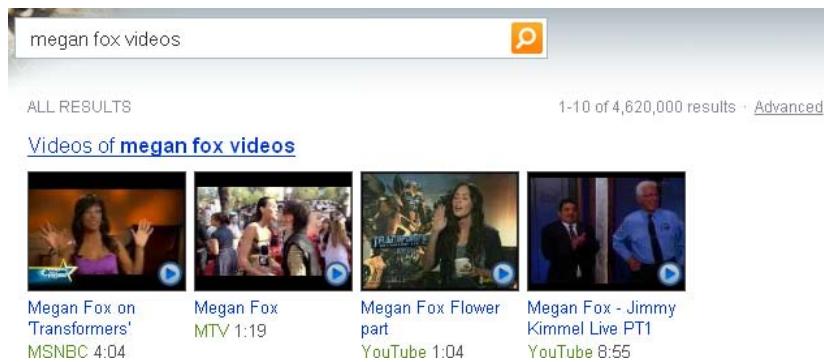


FIGURE 2-10. Bing result for Megan Fox videos

As you can see, the vast variety of vertical integration into search results means that for many popular queries, the standard set of 10 links to external pages is no longer the rule. Engines are competing by providing more relevant results and more targeted responses to queries that they feel are best answered by vertical results, rather than web results.

As a direct consequence, site owners and web marketers must take into account how this incorporation of vertical search results may impact their rankings and traffic. For many of the searches shown in the previous figures, a high ranking—even in position #1 or #2 in the algorithmic/organic results—may not produce much traffic because of the presentation of the vertical results above them.

The vertical results also signify an opportunity, as listings are available in services from images to local search to news and products. We will cover how to get included in these results in Chapter 8.

Algorithm-Based Ranking Systems: Crawling, Indexing, and Ranking

Understanding how crawling, indexing, and ranking works is helpful to SEO practitioners, as it helps them determine what actions to take to meet their goals. This section primarily covers the way Google, Yahoo!, and Microsoft operate, and does not necessarily apply to other search engines that are popular, such as Baidu (China) and Naver (Korea).

The search engines have several major goals and functions. These include:

- Crawling and indexing the billions of documents (pages and files) accessible on the Web
- Responding to user queries by providing lists of relevant pages

In this section, we'll walk through the basics of these functions from a nontechnical perspective. This section will start by discussing how search engines find and discover content.

Crawling and Indexing

Imagine the World Wide Web as a network of stops in a big city subway system. Each stop is its own unique document (usually a web page, but sometimes a PDF, JPEG, or other file). The search engines need a way to “crawl” the entire city and find all the stops along the way, so they use the best path available: the links between web pages, an example of which is shown in [Figure 2-11](#).

In our representation in [Figure 2-11](#), stops such as Embankment, Piccadilly Circus, and Moorgate serve as pages, while the lines connecting them represent the links from those pages to other pages on the Web. Once Google (at the bottom) reaches Embankment, it sees the links pointing to Charing Cross, Westminster, and Temple and can access any of those “pages.”

The link structure of the Web serves to bind together all of the pages that were made public as a result of someone linking to them. Through links, search engines’ automated robots, called *crawlers* or *spiders* (hence the illustrations in [Figure 2-11](#)), can reach the many billions of interconnected documents.

Once the engines find these pages, their next job is to parse the code from them and store selected pieces of the pages in massive arrays of hard drives, to be recalled when needed in a query. To accomplish the monumental task of holding billions of pages that can be accessed in a fraction of a second, the search engines have constructed massive data centers to deal with all this data.

One key concept in building a search engine is deciding where to begin a crawl of the Web. Although you could theoretically start from many different places on the Web, you would ideally begin your crawl with a trusted set of websites. You can think of a factor in evaluating the trust in your website as the click distance between your website and the most trusted

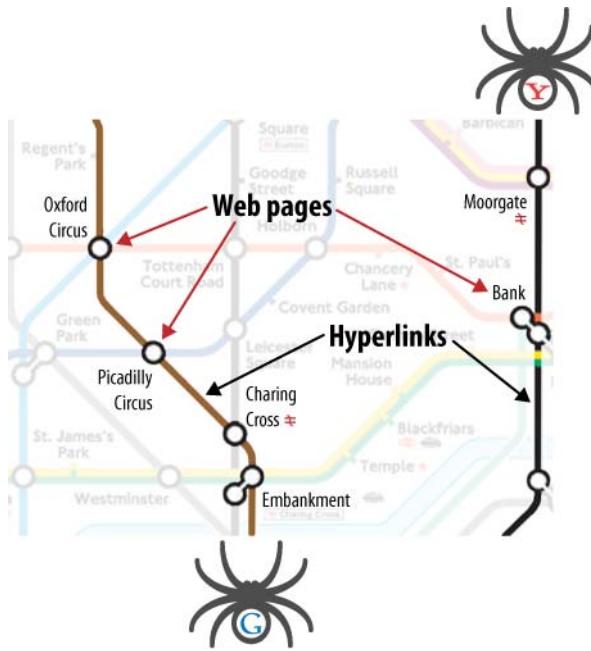


FIGURE 2-11. London's Tube used as an analogy for web crawling

websites. We will discuss the role of trust in search algorithms in more detail in “[How Links Influence Search Engine Rankings](#)” on page 279 in [Chapter 7](#).

Retrieval and Rankings

For most searchers, the quest for knowledge begins as shown in [Figure 2-12](#).



FIGURE 2-12. Start of a user's search quest

The next step in this quest for knowledge occurs when the search engine returns a list of relevant pages on the Web in the order most likely to satisfy the user. This process requires the search engines to scour their corpus of billions of documents and do two things: first, return only the results that are related to the searcher’s query; and second, rank the results in order of perceived importance (taking into account the trust and authority associated with the site). It is both relevance and importance that the process of SEO is meant to influence.

Relevance is the degree to which the content of the documents returned in a search matches the user’s query intention and terms. The relevance of a document increases if the terms or

phrase queried by the user occurs multiple times and shows up in the title of the work or in important headlines or subheads, or if links to the page come from relevant pages and use relevant anchor text.

You can think of relevance as the first step to being “in the game.” If you are not relevant to a query, the search engine does not consider you for inclusion in the search results for that query. We will discuss how relevance is determined in more detail in [“Determining Searcher Intent and Delivering Relevant, Fresh Content” on page 41](#).

Importance or *popularity* refers to the relative importance, measured via *citation* (the act of one work referencing another, as often occurs in academic and business documents) of a given document that matches the user’s query. The popularity of a given document increases with every other document that references it. In the academic world, this concept is known as *citation analysis*.

You can think of importance as a way to determine which page, from a group of equally relevant pages, shows up first in the search results, which is second, and so forth. The relative authority of the site, and the trust the search engine has in it, are significant parts of this determination. Of course, the equation is a bit more complex than this and not all pages are equally relevant. Ultimately, it is the combination of relevance and importance that determines the ranking order.

So, when you see a search results page such as the one shown in [Figure 2-13](#), you can surmise that the search engine (in this case, Yahoo!) believes the Superhero Stamps page on [USPS.com](#) has the highest combined score for relevance and popularity for the query *marvel superhero stamps*, while the Yahoo! Shopping page has a lower combined score for relevance and popularity.

Popularity and relevance aren’t determined manually (those trillions of man-hours would require Earth’s entire population as a workforce). Instead, the engines craft careful, mathematical equations—algorithms—to sort the wheat from the chaff and to then rank the wheat in order of quality. These algorithms often comprise hundreds of components. In the search marketing field, they are often referred to as *ranking factors* or *algorithmic ranking criteria*.

We discuss ranking factors or signals (*signals* is the term Google prefers) in more detail in [“Analyzing Ranking Factors” on page 49](#).

Evaluating Content on a Web Page

Search engines place a lot of weight on the content of each web page. After all, it is this content that defines what a page is about, and the search engines do a detailed analysis of each web page they find during their crawl to help make that determination.

You can think of this as the search engine performing a detailed analysis of all the words and phrases that appear on a web page, and then building a map of that data for it to consider showing your page in the results when a user enters a related search query. This map, often



Also try: [marvel superhero stamps usps](#), [More...](#)

[USPS Stamp News: Spider-Man and Nine Other Marvel Super Heroes](#)

... quality b-roll of the **Marvel** Super Heroes **stamps** printing is available via ... stars from the **Marv** Universe commemorated as **stamps** and stamped postal cards ...
www.usps.com/communications/newsroom/2007/sr07_024.htm - [Cached](#)

[The Ultimate Marvel Collection](#)

Marvel Comics **Stamps** Ultimate **Marvel** Comics Solid Silver **Stamp** & Medal Collection. ...
 Mint-condition sheet of 20 original **Marvel** heroes **stamps** ...
www.ultimatemarvelcollection.com - [Cached](#)

[Marvel Stamps & Stickers - Yahoo! Shopping](#)

Find the best price for **Marvel Stamps** & Stickers at Yahoo! Shopping. Compare prices, read reviews and ratings for **Marvel Stamps** & Stickers
shopping.yahoo.com/s:Stamps & Stickers:4168-Brand=Marvel - 54k

[Comic-Con 2007: Marvel Super Hero Stamp Ceremony | Marvel.com News ...](#)

Check out photos and details from the first panel of San Diego Comic-Con 2007, the unveiling of the **Marvel** USPS commemorative **stamps**!
marvel.com/news/comicstories.1347.Comic-Con_2007:_Marvel_Super_Hero_Stamp... - 92k - {

FIGURE 2-13. Sample search result for “marvel superhero stamps”

referred to as a *semantic map*, seeks to define the relationships between those concepts so that the search engine can better understand how to match the right web pages with user search queries.

If there is no semantic match of the content of a web page to the query, the page has a much lower possibility of showing up. Therefore, the words you put on the page, and the “theme” of that page, play a huge role in ranking.

Figure 2-14 shows how a search engine will break up a page when it looks at it, using a page on the Stone Temple Consulting website.

The navigational elements of a web page are likely similar across the many pages of a site. These navigational elements are not ignored, and they do play an important role, but they do not help a search engine determine what the unique content is on a page. To do that, the search engine gets very focused on the part of Figure 2-14 that is labeled “Real content”.

Determining the unique content on a page is an important part of what the search engine does. It is this understanding of the unique content on a page that the search engine uses to determine the types of search queries for which the web page might be relevant. Since site navigation is generally not unique to a single web page, it does not help the search engine with that task.

Navigation

Eric Enge Interviews Google's Matt Cutts

Published: October 8, 2007

Matt Cutts joined Google as a Software Engineer in January 2000. Before Google, he was working at his Ph.D. in computer graphics at the University of North Carolina at Chapel Hill. He has an M.S. from UNC-Chapel Hill, and B.S. degrees in both mathematics and computer science from the University of Kentucky.

Matt wrote SafeSearch, which is Google's family filter. In addition to his work on Google, Matt had a top-secret clearance while working for the Department of Defense, and he's also worked at a game engine company. He claims that Google is the most fun he's ever had.

Matt currently heads up the Webspam team for Google. Matt talks about webmaster-related issues on his blog.

Interviews

- Google's Peter Linzay, July 12, 2009
- Yahoo's Dennis Montosian, June 15, 2009
- Enquiro's Richard Zwicky, June 15, 2009
- Microsoft's Dr. Scott Prevost
- YouTube's Matt Cutts and John Lu
- 2009 SEMPO John Mueller, May 14, 2009
- SEOmoz's Sean Bird, May 4, 2009
- Overture's Bill Margolis, April 6, 2009
- Google's Rajat Mukherjee, March 31, 2009
- Tatina Parsons, March 23, 2009
- InfoUSA's Parikla Mather, March 17, 2009
- Google's Matt Cutts, March 17, 2009
- Google's Matt Cutts, February 12, 2009
- Google's Bruno Crama, February 1, 2009
- Google's Cedric Dupont, February 4, 2009
- Market Motive's Scott Mirza, Jan. 26, 2009
- AJR's Motoko Hunt, January 12, 2009

FIGURE 2-14. Breaking up a web page

This does not mean navigation links are not important, because they most certainly are—however, they simply do not count when trying to determine the unique content of a web page because those navigation links are shared among many web pages.

One task the search engines face is judging the value of content. Although evaluating how the community responds to a piece of content using link analysis is part of the process, the search engines can also draw some conclusions based on what they see on the page.

For example, is the exact same content available on another website? Is the unique content the search engine can see two sentences long or 500+ words long? Does the content repeat the same keywords excessively? These are a few examples of things that the search engine can look at when trying to determine the value of a piece of content.

What Content Can Search Engines “See” on a Web Page?

Search engine crawlers and indexing programs are basically software programs. These programs are extraordinarily powerful. They crawl hundreds of billions of web pages, analyze the content of all these pages, and analyze the way all these pages link to each other. Then they organize this into a series of databases that can respond to a user search query with a highly tuned set of results in a few tenths of a second.

This is an amazing accomplishment, but it has its limitations. Software is very mechanical, and it can understand only portions of most web pages. The search engine crawler analyzes the raw HTML form of a web page. If you want to see what this looks like, you can do so by using your browser to view the source.

The two screen shots in Figure 2-15 show how to do that in Firefox (top) and Internet Explorer (bottom).



FIGURE 2-15. Viewing source in your browser

Once you view the source, you will be presented with the exact code for the website that the web server sent to your browser. This is what the search engine crawler sees (the search engine also sees the HTTP headers for the page). The crawler will ignore a lot of what is in the code. For example, search engines largely ignore code such as that shown in Figure 2-16, as it has nothing to do with the content of the web page.

```

<div id="wrapper">
  <div id="page-top">
    <div class="container">
      <ul id="site-nav">
        <li id="user_state">Aloha, <a href="/users/view/18040">Eric Enge</a></li>
        <li><a href="/messages" id="messages">0 Messages</a></li>
        <li><a href="/users/my">My Account</a></li>
        <li><a href="/users/logout" onclick="return confirm('Are you sure you want to logout?');">Logout</a>
      </ul>
      <!-- PRO Tour upgrade link in the nav -->
      <li>
        <form id="site-search" action="http://search.yahoo.com/search" method="get">
          <fieldset class="val">
            <input type="hidden" name="vs" value="seomoz.org" />
            <input type="text" name="p" id="site-search-text" tabindex="5" title="Search SEOMoz.org" />
            <input type="submit" name="site-search-go" id="site-search-go" value="SEARCH" />
          </fieldset>
        </form>
      </li>
    </ul>
  </div> <!-- end container -->
</div> <!-- end page top -->
```

FIGURE 2-16. Sample web page source code

The information the search engine crawler is most interested in is in the HTML text on the page. **Figure 2-17** is an example of HTML text for a web page (using the [SEOMoz.org](#) home page).

```

<li class="last">
<em><a href="#" hidefocus="true" id="money1" href="r/wq">Life</a></em>
<span class="pipe"></span>
</li>
</ul>
</div>
<div id="todayvideo">
</div>
<div id="todaypod" class="bd">
<cite class="timestamp">&nbsp;</cite>
<span id="featuredict" class="current">
<div id="featuredmain">
<a href="s/1102897">In this July 20, 1969 file photo, Astronaut Edwin E. "Buzz" Aldrin Jr. poses for a photograph beside the U.S. flag deployed on the moon during the Apollo 11 mission. (AP Photo/Neil Armstrong, NASA, file)"><a>
<span>
<h3><a href="s/1102897">Astronauts' pitch to Obama</a></h3>
<p>The 40th anniversary of man's first steps on the moon sparks a bold request for the president.</p>
<a class="more" href="s/1102897">#187; The future</a></p>
<ul>
<li><a class="search" href="s/1102898">Find habitable planets</a></li>
<li><a class="bullet" href="s/1102899">Lost word</a></li>
<li><a class="video" href="s/1102900" onclick="window.open('s/1102900','playerWindow','width=793,height=666,scrollbars=no');return false;">Video</a>
```

FIGURE 2-17. Sample HTML text in the source code

Although **Figure 2-17** still shows some HTML encoding, you can see the “regular” text clearly in the code. This is the unique content that the crawler is looking to find.

In addition, search engines read a few other elements. One of these is the page title. The page title is one of the most important factors in ranking a given web page. It is the text that shows in the browser’s title bar (the blue line above the browser menu and the address bar).

Figure 2-18 shows the code that the crawler sees, using [Trip Advisor](#) as an example.

The first red circle in **Figure 2-18** is for the title tag. The title tag is also often (but not always) used as the title of your listing in search engine results. Exceptions to this can occur when you obtain Yahoo! or DMOZ directory listings for your site. Sometimes the search engines may

choose to use a title for your page that was used in your listings in these directories, instead of the title tag on the page. There are also meta tags that allow you to block this from happening, such as the NOODP tag, which tells the search engine not to use DMOZ titles, and the NOYDIR tag, which tells Yahoo! not to use the Yahoo! directory listing. In any event, Figure 2-19 shows what happens when you search on *stone temple consulting* (the Stone Temple Consulting home page at <http://www.stonetemple.com>). Notice how the title of the search listing matches the title of the Stone Temple Consulting home page.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
<meta http-equiv="imagetoolbar" content="no"/>
<title>Reviews of vacations, hotels, resorts, vacation and travel packages - TripAdvisor</title>
<meta http-equiv="pragma" content="no-cache"/>
<meta http-equiv="cache-control" content="no-cache,must-revalidate"/>
<meta http-equiv="Expires" content="0"/>
<meta name="Keywords" content="vacation, vacations, vacation packages, vacation package, travel package,
travel packages, travel, planning, hotel, hotels, motel, bed and breakfast, inn, guidebook, review, reviews,
popular, plan, airfare, cheap, discount, map, maps, golf, ski, articles, attractions, advice, restaurants"/>
<meta name="Description" content="TripAdvisor - Unbiased hotel reviews, photos and travel advice for hotels
and vacations. Compare prices with just one click."/>
<script type="text/javascript">
var pageInit = new Date();
var jsGlobalMonths = new
```

FIGURE 2-18. Meta tags in HTML source

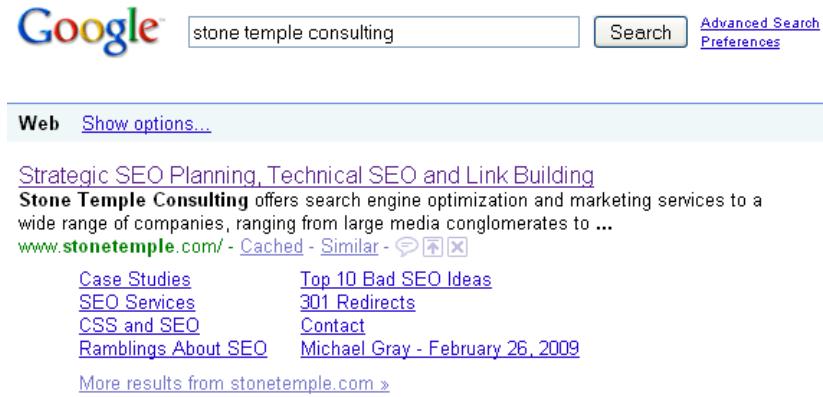


FIGURE 2-19. Search result showing title tag

In addition to page titles, the search engines also read the “meta keywords” tag. This is a list of keywords that you wish to have associated with the page. *Spammers* (people who attempt to manipulate search engine results in violation of the search engine guidelines) ruined the SEO value of this tag many years ago, so its value is now negligible. Google does not use this tag for ranking at all, but Yahoo! and Bing seem to make reference to it (you can read about this in detail at <http://searchengineland.com/meta-keywords-tag-101-how-to-legally-hide-words-on-your-pages-for-search-engines-12099>). Spending a lot of time on meta keywords is not recommended because of the lack of SEO benefit.

The second red circle in [Figure 2-18](#) shows an example of a meta keywords tag.

Search engines also read the meta description tag (the third red circle in the HTML in [Figure 2-18](#)). However, the meta description tag is not of any influence in search engine rankings (<http://searchengineland.com/21-essential-seo-tips-techniques-11580>).

Nonetheless, the meta description tag plays a key role as search engines often use it as the description for your page in search results. Therefore, a well-written meta description can have a significant influence on how many clicks you get on your search listing. Time spent on meta descriptions is quite valuable as a result. [Figure 2-20](#) uses a search on *trip advisor* to show an example of the meta description being used as a description in the search results.

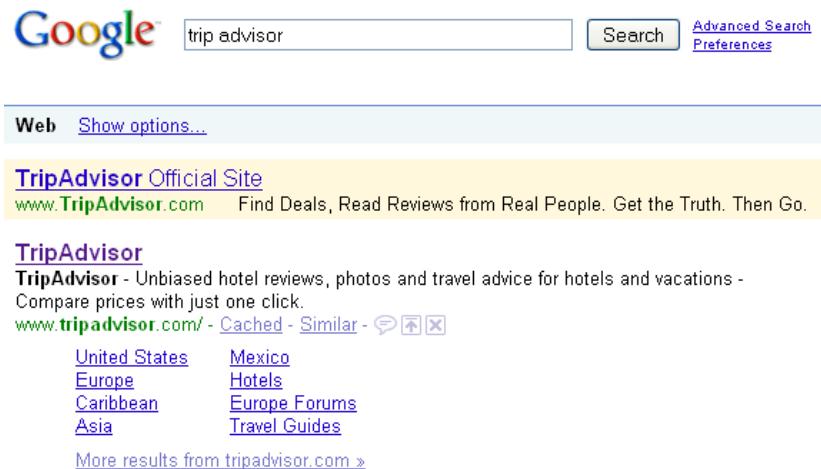


FIGURE 2-20. Meta description used in search results

NOTE

The user's keywords are typically shown in boldface when they appear in the search results (sometimes close synonyms are shown in boldface as well). As an example of this, in [Figure 2-20](#), *TripAdvisor* is in boldface at the beginning of the description.

A fourth element that search engines read is the alt attribute for images. The alt attribute was originally intended to allow something to be rendered when viewing of the image is not possible. There were two basic audiences for this:

- Vision-impaired people who do not have the option of viewing the images.
- People who turn off images for faster surfing. This is generally an issue only for those who do not have a broadband connection.

Support for the vision-impaired remains a major reason for using the alt attribute. You can read more about this by visiting the Web Accessibility Initiative page on the [W3C website](#).

Search engines also read the text contained in the alt attribute of an image tag. An image tag is an element that is used to tell a web page to display an image. Here is an example of an image tag from the Alchemist Media site:

```

```

The src= part of the tag is where the image to be displayed is located. The part that starts with alt= and is then followed with "BtoB Interactive Marketing Guide" is considered the alt attribute.

The alt attribute is something that the search engines read. The search engines interpret it to help them determine what the image is about and to get more of a sense as to what the page is about.

A final element that search engines read is the noscript tag. In general, search engines do not try to interpret JavaScript that may be present on a web page (though this is already changing). However, a small percentage of users do not allow JavaScript to run when they load a web page (the authors' experience is that it is about 2%). For those users, nothing would be shown where the JavaScript is on the web page, unless the page contains a noscript tag.

Here is a very simple JavaScript example that demonstrates this:

```
<script type="text/javascript">
document.write("It is a Small World After All!")
</script>
<noscript>Your browser does not support JavaScript!</noscript>
```

The noscript portion of this is Your browser does not support JavaScript!. The search engines will read this text and see that as information about the web page. In this example, you could also choose to make the noscript tag contain the text "It is a Small World After All!". The noscript tag should be used only to represent the content of the JavaScript. (Placing other content or links in this tag could be interpreted as spammy behavior by the search engines.) In addition, the browser warning could end up as your search snippet, which would be a bad thing.

What search engines cannot see

It is also worthwhile to review the types of content that search engines cannot “see” in the human sense.

For instance, although search engines are able to detect that you are displaying an image, they have little idea what the image is a picture of, except for whatever information you provide them in the alt attribute, as discussed earlier. They can, however, determine pixel color and, in many instances, determine whether images have pornographic content by how much flesh tone there is in a JPEG image. So, a search engine cannot tell whether an image is a picture of Bart Simpson, a boat, a house, or a tornado. In addition, search engines will not recognize any text rendered in the image. The search engines are experimenting with technologies to use optical character recognition (OCR) to extract text from images, but this technology is not yet in general use within search.

In addition, conventional SEO wisdom has always held that the search engines cannot read Flash files, but this is a little overstated. Search engines are beginning to extract information from Flash, as indicated by the Google announcement at <http://googlewebmastercentral.blogspot.com/2008/06/improved-flash-indexing.html>. However, the bottom line is that it's not easy for search engines to determine what is in Flash. One of the big issues is that even when search engines look inside Flash, they are still looking for textual content, but Flash is a pictorial medium and there is little incentive (other than the search engines) for a designer to implement text inside Flash. All the semantic clues that would be present in HTML text (such as heading tags, boldface text, etc.) are missing too, even when HTML is used in conjunction with Flash.

A third type of content that search engines cannot see is the pictorial aspects of anything contained in Flash, so this aspect of Flash behaves in the same way images do. For example, when text is converted into a vector-based outline (i.e., rendered graphically), the textual information that search engines can read is lost. We discuss methods for optimizing Flash in “[Redirects](#)” on page 190 in [Chapter 6](#).

Audio and video files are also not easy for search engines to read. As with images, the data is not easy to parse. There are a few exceptions where the search engines can extract some limited data, such as ID3 tags within MP3 files, or enhanced podcasts in AAC format with textual “show notes,” images, and chapter markers embedded. Ultimately, though, a video of a soccer game cannot be distinguished from a video of a forest fire.

Search engines also cannot read any content contained within a program. The search engine really needs to find text that is readable by human eyes looking at the source code of a web page, as outlined earlier. It does not help if you can see it when the browser loads a web page—it has to be visible and readable in the source code for that page.

One example of a technology that can present significant human-readable content that the search engines cannot see is AJAX. AJAX is a JavaScript-based method for dynamically rendering content on a web page after retrieving the data from a database, without having to

refresh the entire page. This is often used in tools where a visitor to a site can provide some input and the AJAX tool then retrieves and renders the correct content.

The problem arises because the content is retrieved by a script running on the client computer (the user's machine) only after receiving some input from the user. This can result in many potentially different outputs. In addition, until that input is received the content is not present in the HTML of the page, so the search engines cannot see it.

Similar problems arise with other forms of JavaScript that don't render the content in the HTML until a user action is taken.

As of HTML 5, a construct known as the embed tag (`<embed>`) was created to allow the incorporation of plug-ins into an HTML page. Plug-ins are programs located on the user's computer, not on the web server of your website. This tag is often used to incorporate movies or audio files into a web page. The `<embed>` tag tells the plug-in where it should look to find the datafile to use. Content included through plug-ins is not visible at all to search engines.

Frames and iframes are methods for incorporating the content from another web page into your web page. Iframes are more commonly used than frames to incorporate content from another website. You can execute an iframe quite simply with code that looks like this:

```
<iframe src ="http://accounting.careerbuilder.com" width="100%" height="300">
  <p>Your browser does not support iframes.</p>
</iframe>
```

Frames are typically used to subdivide the content of a publisher's website, but they can be used to bring in content from other websites, as was done in [Figure 2-21](#) with <http://accounting.careerbuilder.com> on the *Chicago Tribune* website.

[Figure 2-21](#) is an example of something that works well to pull in content (provided you have permission to do so) from another site and place it on your own. However, the search engines recognize an iframe or a frame used to pull in another site's content for what it is, and therefore ignore the content inside the iframe or frame as it is content published by another publisher. In other words, they don't consider content pulled in from another site as part of the unique content of your web page.

Determining Searcher Intent and Delivering Relevant, Fresh Content

Modern commercial search engines rely on the science of information retrieval (IR). This science has existed since the middle of the twentieth century, when retrieval systems powered computers in libraries, research facilities, and government labs. Early in the development of search systems, IR scientists realized that two critical components comprised the majority of search functionality: relevance and importance (which we defined earlier in this chapter). To measure these factors, search engines perform document analysis (including semantic analysis of concepts across documents) and link (or citation) analysis.

Keywords: e.g. Manager or Sales or enter a Web ID **Location:** Chicago, IL **Posted within:** Last 30 Days **Find Jobs**

Narrow Search

Category:	Company:	City:	State:
Accounting (757)	Accounting Now (47)	Chicago (382)	Illinois (785)
Finance (523)	Jackson Wabash (38)	Schaumburg (34)	Indiana (3)
Management (190)	Robert Half Finance & ... (33)	Oak Brook (27)	
Customer Service (86)	Ajilon Professional... (29)	Naperville (24)	
Admin - Clerical (74)	SNI Financial (23)	Hoffman Estates (19)	

788 ACCOUNTING JOBS FOUND WITHIN 30 MILES OF CHICAGO, IL

Show me: Closest jobs first | Relevant jobs first | Newest jobs first

Search Results

Job Title / Description (show titles only)	Company	Location (Distance)	Posted
MANAGER, COMMISSIONS View similar jobs Job type: Full-Time Employee Follett Higher Education We help institutions, administrators, faculty and students achieve a shared goal: to fully realize the powerful benefits of... View full job description Save to MyCareerBuilder Email to a friend	Follett Higher Education Group	IL - Oak Brook (16 miles)	Today
Payroll Administrator View similar jobs Job type: Full-Time Employee Pay: \$19.00 - \$21.63/hour Payroll Administrator We are looking for an experienced Payroll Administrator to join a growing company in Oak Brook, IL with a great work environment... View full job description Save to MyCareerBuilder Email to a friend	Torrey & Gray	120 W. Madison St. IL - Chicago (2 miles)	Today
Payroll Administrator View similar jobs Job type: Full-Time Employee Insurance Auto Auctions, Inc. seeks a Payroll Administrator to assist in processing our bi-weekly payroll for approximately 1800 employees nationwide... View full job description Save to MyCareerBuilder Email to a friend	Insurance Auto Auctions, Inc.	Two Westbrook Corporate Center IL - Westchester (13 miles)	Today

FIGURE 2-21. Framed page rendered in a browser

Document Analysis and Semantic Connectivity

In *document analysis*, search engines look at whether they find the search terms in important areas of the document—the title, the metadata, the heading tags, and the body of the text. They also attempt to automatically measure the quality of the document based on document analysis, as well as many other factors.

Reliance on document analysis alone is not enough for today's search engines, so they also look at semantic connectivity. *Semantic connectivity* refers to words or phrases that are commonly associated with one another. For example, if you see the word *aloha* you associate it with Hawaii, not Florida. Search engines actively build their own thesaurus and dictionary to help them determine how certain terms and topics are related. By simply scanning their massive databases of content on the Web, they can use Fuzzy Set Theory and certain equations

(described at <http://forums.searchenginewatch.com/showthread.php?threadid=48>) to connect terms and start to understand web pages/sites more like a human does.

The professional SEO practitioner does not necessarily need to use semantic connectivity measurement tools to optimize websites, but for those advanced practitioners who seek every advantage, semantic connectivity measurements can help in each of the following sectors:

- Measuring which keyword phrases to target
- Measuring which keyword phrases to include on a page about a certain topic
- Measuring the relationships of text on other high-ranking sites/pages
- Finding pages that provide “relevant” themed links

Although the source for this material is highly technical, SEO specialists need only know the principles to obtain valuable information. It is important to keep in mind that although the world of IR has hundreds of technical and often difficult-to-comprehend terms, these can be broken down and understood even by an SEO novice.

Table 2-1 explains some common types of searches in the IR field.

TABLE 2-1. Common types of searches

Proximity searches	A proximity search uses the order of the search phrase to find related documents. For example, when you search for “ <i>sweet German mustard</i> ” you are specifying only a precise proximity match. If the quotes are removed, the proximity of the search terms still matters to the search engine, but it will now show documents that don’t exactly match the order of the search phrase, such as <i>Sweet Mustard—German</i> .
Fuzzy logic	Fuzzy logic technically refers to logic that is not categorically true or false. A common example is whether a day is sunny (is 50% cloud cover a sunny day, etc.). In search, fuzzy logic is often used for misspellings.
Boolean searches	These are searches that use Boolean terms such as AND, OR, and NOT. This type of logic is used to expand or restrict which documents are returned in a search.
Term weighting	Term weighting refers to the importance of a particular search term to the query. The idea is to weight particular terms more heavily than others to produce superior search results. For example, the appearance of the word <i>the</i> in a query will receive very little weight in selecting the results because it appears in nearly all English language documents. There is nothing unique about it, and it does not help in document selection.

IR models (search engines) use Fuzzy Set Theory (an offshoot of fuzzy logic created by Dr. Lotfi Zadeh in 1969) to discover the semantic connectivity between two words. Rather than using a thesaurus or dictionary to try to reason whether two words are related to each other, an IR system can use its massive database of content to puzzle out the relationships.

Although this process may sound complicated, the foundations are simple. Search engines need to rely on machine logic (true/false, yes/no, etc.). Machine logic has some advantages over humans, but machine logic doesn’t have a way of thinking like humans, and things that are

intuitive to humans can be quite hard for a computer to understand. For example, both oranges and bananas are fruits, but both oranges and bananas are not round. To a human this is intuitive.

For a machine to understand this concept and pick up on others like it, semantic connectivity can be the key. The massive human knowledge on the Web can be captured in the system's index and analyzed to artificially create the relationships humans have made. Thus, a machine knows an orange is round and a banana is not by scanning thousands of occurrences of the words *banana* and *orange* in its index and noting that *round* and *banana* do not have great concurrence, while *orange* and *round* do.

This is how the use of fuzzy logic comes into play, and the use of Fuzzy Set Theory helps the computer to understand how terms are related simply by measuring how often and in what context they are used together.

A related concept that expands on this notion is latent semantic analysis (LSA). The idea behind this is that by taking a huge composite (index) of billions of web pages, the search engines can "learn" which words are related and which noun concepts relate to one another.

For example, using LSA, a search engine would recognize that *trips* to the *zoo* often include *viewing wildlife* and *animals*, possibly as part of a *tour*.

Now, conduct a search on Google for *~zoo ~trips* (the tilde is a search operator; more on this later in this chapter). Note that the boldface words that are returned match the terms that are italicized in the preceding paragraph. Google is setting "related" terms in boldface and recognizing which terms frequently occur concurrently (together, on the same page, or in close proximity) in their indexes.

Some forms of LSA are too computationally expensive. For example, currently the search engines are not smart enough to "learn" the way some of the newer learning computers do at MIT. They cannot, for example, learn through their index that *zebras* and *tigers* are examples of striped animals, although they may realize that *stripes* and *zebras* are more semantically connected than *stripes* and *ducks*.

Latent semantic indexing (LSI) takes this a step further by using semantic analysis to identify related web pages. For example, the search engine may notice one page that talks about doctors and another one that talks about physicians, and determine that there is a relationship between the pages based on the other words in common between the pages. As a result, the page referring to doctors may still show up for a search query that uses the word *physician* instead.

Search engines have been investing in these types of technologies for many years. For example, in April 2003 Google acquired [Applied Semantics](#), a company known for its semantic-text-processing technology. This technology currently powers Google's AdSense advertising program, and has most likely made its way into the core search algorithms as well.

For SEO purposes, this usage opens our eyes to realizing how search engines recognize the connections between words, phrases, and ideas on the Web. As semantic connectivity becomes

a bigger part of search engine algorithms, you can expect greater emphasis on the theme of pages, sites, and links. It will be important going into the future to realize the search engines' ability to pick up on ideas and themes and recognize content, links, and pages that don't fit well into the scheme of a website.

Link Analysis

In link analysis, search engines measure who is linking to a site or page and what they are saying about that site/page. They also have a good grasp on who is affiliated with whom (through historical link data, the site's registration records, and other sources), who is worthy of being trusted based on the authority of sites linking to them, and contextual data about the site on which the page is hosted (who links to that site, what they say about the site, etc.).

Link analysis goes much deeper than counting the number of links a web page or website has, as links are not created equal. Links from a highly authoritative page on a highly authoritative site will count more than other links of lesser authority. A website or page can be determined to be authoritative by combining an analysis of the linking patterns and semantic analysis.

For example, perhaps you are interested in sites about dog grooming. Search engines can use semantic analysis to identify the collection of web pages that focus on the topic of dog grooming. The search engines can then determine which of these sites about dog grooming have the most links from the set of dog grooming sites. These sites are most likely more authoritative on the topic than the others.

The actual analysis is a bit more complicated than that. For example, imagine that there are five sites about dog grooming with a lot of links from pages across the Web on the topic, as follows:

- Site A has 213 topically related links.
- Site B has 192 topically related links.
- Site C has 203 topically related links.
- Site D has 113 topically related links.
- Site E has 122 topically related links.

Further, it may be that Site A, Site B, Site D, and Site E all link to each other, but none of them link to Site C. In fact, Site C appears to have the great majority of its relevant links from other pages that are topically relevant but have few links to them. In this scenario, Site C is definitively not authoritative because it is not linked to by the right sites.

This concept of grouping sites based on their relevance is referred to as a *link neighborhood*. The neighborhood you are in says something about the subject matter of your site, and the number and quality of the links you get from sites in that neighborhood say something about how important your site is to that topic.

The degree to which search engines rely on evaluating link neighborhoods is not clear, and is likely to differ among search engines. In addition, links from nonrelevant pages are still believed to help the rankings of the target pages. Nonetheless, the basic idea remains that a link from a relevant site should count for more than a link from a nonrelevant site.

Another factor in determining the value of a link is the way the link is implemented and where it is placed. For example, the text used in the link itself (i.e., the actual text that will go to your web page when the user clicks on it) is also a strong signal to the search engines.

This is referred to as *anchor text*, and if that text is keyword-rich (with keywords relevant to your targeted search terms), it will do more for your rankings in the search engines than if the link is not keyword-rich. For example, anchor text of “Dog Grooming Salon” will bring more value to a dog grooming salon’s website than anchor text of “Click here”. However, take care. If you get 10,000 links using the anchor text “Dog Grooming Salon” and you have few other links to your site, this definitely does not look natural and could lead to problems in your rankings.

The semantic analysis of a link’s value does go deeper than just the anchor text. For example, if you have that “Dog Grooming Salon” anchor text on a web page that is not really about dog grooming at all, the value of the link is less than if the page is about dog grooming. Search engines also look at the content on the page immediately surrounding the link, as well as the overall context and authority of the website that is providing the link.

All of these factors are components of link analysis, which we will discuss in greater detail in [Chapter 7](#).

Problem Words, Disambiguation, and Diversity

On the opposite side of the coin are words that present an ongoing challenge for the search engines. One of the greatest challenges comes in the form of disambiguation. For example, when someone types in *boxers*, does he mean the prize fighter, the breed of dog, or the type of underwear? Another example of this is *jaguar*, which is at once a jungle cat, a car, a football team, an operating system, and a guitar. Which does the user mean?

Search engines deal with these types of ambiguous queries all the time. The two examples offered here have inherent problems built into them, but the problem is much bigger than that. For example, if someone types in a query such as *cars*, does he:

- Want to read reviews?
- Want to go to a car show?
- Want to buy one?
- Want to read about new car technologies?

The query *cars* is so general that there is no real way to get to the bottom of the searcher’s intent based on this one query alone. (The exception is cases where prior queries by the same

searcher may provide additional clues that the search engine can use to better determine the searcher's intent.)

This is the reason the search engines offer diverse results. As an example, Figure 2-22 shows another generic search, this time using *gdp*.



FIGURE 2-22. Diverse results example

This brings up an important ranking concept. It is possible that a strict analysis of the relevance and link popularity scores in Figure 2-22 would not have resulted by itself in the Investopedia.com page being on the first page, *but* the need for diversity caused the ranking of the page to be elevated.

A strict relevance- and importance-based ranking system might have shown a variety of additional government pages discussing the GDP of the United States. However, a large percentage of users will likely be satisfied by the government pages already shown, and showing more of them is not likely to raise the level of satisfaction with the results.

Introducing a bit of variety allows Google to also provide a satisfactory answer to those who are looking for something different from the government pages. Google's testing has shown that this diversity-based approach has resulted in a higher level of satisfaction among its users.

For example, the testing data for the nondiversified results may have shown lower click-through rates in the SERPs, greater numbers of query refinements, and even a high percentage of related searches performed subsequently.

When Google wants to get really serious about disambiguation, it goes a different route. Check out the SERPs in Figure 2-23.



FIGURE 2-23. Disambiguating search queries

These “horizontal line,” disambiguation-style results appear on many searches where Google thinks the searcher is probably seeking something that his query isn’t producing. They’re especially likely to appear for very general search phrases.

The idea to deliberately introduce diversity into the result algorithm makes sense and can enhance searcher satisfaction for queries such as:

- Company names (where searchers might want to get positive and negative press, as well as official company domains)
- Product searches (where e-commerce-style results might ordinarily fill up the SERPs, but Google tries to provide some reviews and noncommercial, relevant content)
- News and political searches (where it might be prudent to display “all sides” of an issue, rather than just the left- or right-wing blogs that did the best job of baiting links)

Where freshness matters

Much of the time, it makes sense for the search engines to deliver results from older sources that have stood the test of time. However, other times the response should be from newer sources of information.

For example, when there is breaking news, such as an earthquake, the search engines begin to receive queries within seconds, and the first articles begin to appear on the Web within 15 minutes.

In these types of scenarios, there is a need to discover and index new information in near-real time. Google refers to this concept as *query deserves freshness* (QDF). According to the *New York Times* (<http://www.nytimes.com/2007/06/03/business/yourmoney/03google.html?pagewanted=3>), QDF takes several factors into account, such as:

- Search volume
- News coverage
- Blog coverage
- Toolbar data (maybe)

QDF applies to up-to-the-minute news coverage, but also to other scenarios such as hot, new discount deals or new product releases that get strong search volume and media coverage.

A Few Reasons Why These Algorithms Sometimes Fail

As we've outlined in this chapter, the search engines do some amazing stuff. Nonetheless, there are times when the process does not work as well as you would like to think. Part of this is because users often type in search phrases that provide very little information about their intent (e.g., if they search on *car*, do they want to buy one, read reviews, learn how to drive one, learn how to design one, or something else?). Another reason is that some words have multiple meanings, such as *jaguar* (which is an animal, a car, a guitar, and in its plural form, a football team).

For more information on reasons why search algorithms sometimes fail, you can read the following SEOMoz article, which was written by Hamlet Batista:

<http://www.seomoz.org/blog/7-reasons-why-search-engines-dont-return-relevant-results-100-of-the-time>

Analyzing Ranking Factors

SEOMoz conducted a survey of leading SEOs to determine what these SEOs thought were the most important ranking factors (<http://www.seomoz.org/article/search-ranking-factors>). Here is a high-level summary of the top nine results:

- Keyword use in title tag
- Anchor text of inbound link
- Global link authority of site
- Age of site
- Link popularity within the site's internal link structure
- Topical relevance of inbound links
- Link popularity of site in topical community
- Keyword use in body text
- Global link popularity of sites that link to the site

Here is a brief look at each of these:

Keyword use in title tag

The title tag was clearly considered the #1 ranking factor. The words in the title tag say a lot about the topic of a page (or they should), and it is a useful element for the search engines to key on. One reason is that it is a highly visible element on the web page and therefore is likely to be used, well, as a title for the page.

It is also generally accepted that search engines place the most weight on the words that appear at the start of the title. For that reason it is always a good idea to put your most important keywords at the start of the title tag, but make sure to avoid *keyword stuffing* (a concept that we will discuss more in “[Keyword Targeting](#)” on page 211 in [Chapter 6](#)).

Anchor text of inbound link

When one web page links to another, the anchor text used for that link is also used as a strong signal by the search engines. Webmasters and publishers who choose to link to other web pages from their site often will place keyword-rich anchor text in those links.

This has led to a tremendous focus by link builders in an attempt to get websites to provide keyword-rich text links to their clients’ websites.

Global link authority of site

This metric factors the total link authority of the site, as opposed to the link authority of a given page on the site. As defined in the SEOmoz survey, it is intended to capture the quantity and quality of a site’s links. There are three quick ways to get a general sense of this:

- Look at the PageRank for the home page of the domain. Although the original PageRank algorithm is flawed, it can still provide a general sense of a site’s global link popularity (one area where this is weak is that it will not give you a sense of the relevance of any of the links). Note, however, that the home page is not always the highest PageRank page on a site.
- Use the [Yahoo! Site Explorer](#) search. This will give you a raw count of all inbound links to the domain and allow you to download up to 1,000 results into a spreadsheet.
- Use [Linkscape’s](#) domain mozRank.

Age of site

This metric often surprises novice SEO practitioners, but there is no question that it is a factor. The reason this makes for a valuable signal is that a site that has been around for a long time must be doing something right in the eyes of the community it is trying to serve. Spamy or poor-quality sites ultimately don’t do as well, so they don’t last as long. In addition, spam was less prevalent in 1998, so sites started in 1998 are less likely to be spam compared to sites started more recently.

Link popularity within the site's internal link structure

Search engines rely on the publisher to tell them what they think is the most important part of their site. For example, a page that can be reached only after five clicks from the home page is probably not important.

On the other hand, a page that is included in the global navigation of the site, and hence ends up being linked to by every page of the site, is clearly one of the more important pages. The relevance of those internal links is also a factor (the more relevant they are the better).

Topical relevance of inbound links

The relevance of the sites and pages linking to the target page and the target keyword is a big factor as well. Links from a site on an unrelated topic are still beneficial, but highly relevant links to a site count for a lot more.

Link popularity of site in topical community

This metric speaks to establishing the authority of your site. If you have many links from members of the topical community, that is a very strong vote of confidence for your site.

Keyword use in body text

Search engines also look at the text within the body of a page. Putting the relevant keyword in your title tag is good, but it adds to the impact if the on-page content backs it up. Don't limit the content to the exact target keyword; the search engines also look for synonyms to further reinforce the relevance of a page to a topic.

Global link popularity of sites that link to the site

Another ranking factor is the importance of the linking site. Is the site a trusted authority in its field? If so, this will probably show up in its Global Link Popularity.

Negative Ranking Factors

The SEOMoz survey also identified the top five negative factors:

Server is often inaccessible to crawlers

Search engines want their users to have good experiences. If your site is subject to frequent outages, by definition it is not providing a good user experience. So, if the search engine crawler frequently is unable to access your web pages, the search engine will assume that it is dealing with a low-quality site.

Content very similar to or duplicate of other web pages

Search engines want to present lots of unique content (reference our discussion of diversity in ["Determining Searcher Intent and Delivering Relevant, Fresh Content" on page 41](#)). If your website is very thin on content, or is largely a duplicate of the content on other websites, this will potentially be seen as a negative ranking signal.

We discuss duplicate content in more detail in ["Duplicate Content Issues" on page 226](#) in [Chapter 6](#).

External links to low-quality/spam sites

One way to hurt your site's ranking is to make a practice of linking to low-quality sites or spam sites. Generally speaking, one bad outbound link might not hurt you, but if you link to lots of poor-quality sites, it certainly can hurt your rankings.

What makes this more complex is that you might link to a perfectly good domain that goes out of business. Whoever picks up the domain next might not be someone you want to link to, but you may not know that the site you originally linked to is gone and has been replaced by a spam site.

Participation in link schemes or actively selling links

Search engines also do not want publishers to sell links on their site for the purposes of passing *link authority* (or *link juice*, as many people in the industry call it). The search engines don't want publishers participating in shady link schemes, such as distributing hit counters with hidden links within them.

Duplicate titles/meta tags on many pages

Since the title tag is a very powerful signal, failure to create a distinct title tag can be a negative ranking signal. Correspondingly, the meta description, and meta keywords tags duplicated en masse, could potentially be seen as a negative ranking factor.

Other Ranking Factors

The ranking factors we've discussed so far are really just the basics. Search engines potentially factor in many more signals. Some of these include:

Rate of acquisition of links

If over time your site has acquired an average of 5 links per day, and then the links suddenly start to come in at a rate of 10 per day, that could be seen as a positive ranking signal.

On the other hand, if the rate of new links drops to two per day, that could be a signal that your site has become less relevant. Additionally, if your site suddenly starts to get 30 new links per day, you have either become a lot more relevant or started to acquire links in a spammy way. The devil is in the details here, with one of the most important details being from where those new links are coming. The concept of considering temporal factors in link analysis is documented in a U.S. patent held by Google and found at the following URL:

<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.html&r=1&f=G&l=50&s1=%2220050071741%22.PGNR.&OS=DN/20050071741&RS=DN/20050071741>

Alternatively, you can look it up by searching for patent number 20050071741.

Usage data

Other factors can be interpreted as a ranking signal. For example, if a user clicks through on a SERP listing and comes to your site, clicks the Back button, and then clicks on another

result in the same set of search results, that could be seen as a negative ranking signal. Or if the results below you in the SERPs are getting clicked on more than you are, that could be seen as a negative ranking signal for you and a positive ranking signal for them.

Whether search engines use this signal or not, or how much weight they might put on it, is not known. At best, this is a noisy ranking signal and any use is likely to be small in scope. Any use may be primarily related to personalization and relevance feedback.

User data

Personalization is one of the most talked about frontiers in search. There are a few ways personalization can take place. For one, a search engine can perform a geolocation lookup to figure out where you are approximately located. Based on this, a search engine can show results tailored to your current location. This is very helpful, for example, if you are looking for a local restaurant.

Another way a search engine can get some data on a user is if the user creates a profile with the search engine and voluntarily provides some information. A simple example would be a language preference. If the user indicates he prefers Portuguese, the search engine can tailor the results to that preference.

Search engines can also look at the search history for a given user. Basically, the search engine maintains a log of all the searches you have performed when you are logged in. Based on this, it can see that you have been checking out luxury cars recently, and can use that knowledge to tweak the results you see after you search on *jaguar*.

Google sandbox

As we have discussed throughout this chapter, the search engines use a number of methods to fight spam. One technique that many people believe Google uses became known as the Google “sandbox.” The sandbox is thought to be a filter where Google limits the rate of growth of the PageRank (or rankings) of new domains.

This approach could be useful in filtering out spam domains because they often don’t stay around very long, so the spammer works hard to get them ranking and producing traffic as quickly as they can. The sandbox can potentially create a scenario where the website is caught by improved algorithms or manual review prior to becoming highly productive. At a minimum, it would increase the cost of the spammer’s efforts.

Using Advanced Search Techniques

One of the basic tools of the trade for an SEO practitioner is the search engines themselves. They provide a rich array of commands that can be used to perform advanced research, diagnosis, and competitive analysis. Some of the more basic operators are:

[-keyword]

Excludes the keyword from the search results. For example, [loans -student] shows results for all types of loans *except* student loans.

[+keyword]

Allows for forcing the inclusion of a keyword. This is particularly useful for including *stopwords* (keywords that are normally stripped from a search query because they usually do not add value, such as the word *the*) in a query, or if your keyword is getting converted into multiple keywords through automatic stemming. For example, if you mean to search for the TV show *The Office*, you would want the word *The* to be part of the query. As another example, if you are looking for Patrick Powers, who was from Ireland, you would search for *patrick +powers Ireland* to avoid irrelevant results for Patrick Powers.

["key phrase"]

Shows search results for the exact phrase—for example, ["seo company"].

[keyword1 OR keyword2]

Shows results for *at least one* of the keywords—for example, [google OR Yahoo!].

These are the basics, but for those who want more information, what follows is an outline of the more advanced search operators available from the search engines.

Advanced Google Search Operators

Google supports a number of advanced search operators (<http://www.google.com/help/operators.html>) that you can use to help diagnose SEO issues. Table 2-2 gives a brief overview of the queries, how you can use them for SEO purposes, and examples of usage.

TABLE 2-2. Google advanced search operators

Operator	Short description	SEO application	Examples
(site:)—Domain-restricted search	Narrows a search to (a) specific domain(s)/directories	Shows approximately how many URLs are indexed by Google	site:www.google.com
		From a directory	site:mit.edu/research/
		Includes all subdomains	site:google.com
	Shows sites of a specific top-level domain (TLD)	site:org	
(inurl:)/(allinurl:)—URL keyword restricted search	Narrows the results to documents containing (a) search term(s) in the URLs	Find web pages having your keyword in a file path	inurl:seo inurl:company = allinurl:seo company

Operator	Short description	SEO application	Examples
(intitle:)/ (allintitle:)—Title keyword restricted search	Restricts the results to documents containing (a) search term(s) in a page title	Find web pages using your keyword in a page title	intitle:seo intitle:company = allintitle:seo company
(inanchor:)/ (allinanchor:)—Anchor text keyword restricted search	Restricts the results to documents containing (a) search term(s) in the anchor text of backlinks pointing to a page	Find pages having the most backlinks/the most powerful backlinks with the keyword in the anchor text	inanchor:seo inanchor:company = allinanchor:seo company
(intext:)—Body text keyword restricted search	Restricts the results to documents containing (a) search term(s) in the body text of a page	Find pages containing the most relevant/most optimized body text	intext:seo
(ext:)/(filetype:)—File type restricted search	Narrows search results to the pages that end in a particular file extension	A few possible extensions/file types: PDF (Adobe Portable Document Format) HTML or .htm (Hypertext Markup Language) .xls (Microsoft Excel) .ppt (Microsoft PowerPoint) .doc (Microsoft Word)	filetype:pdf ext:pdf
(*)—Wildcard search	Means “insert any word here”	Search for a phrase “partial match”	seo* directory returns “seo free directory”, “seofriendly directory”, etc.
(Related:)—Similar URLs search	Shows <i>related pages</i> by finding pages linking to	Evaluate how relevant the site’s “neighbors” are	Compare: related:www.linkdiy.com

Operator	Short description	SEO application	Examples
	the site and looking at what else they tend to link to (i.e., “co-citation”); usually 25 to 31 results are shown.		and related: www.alchemistmedia.com
(Info:)—Information about a URL search	Gives information about the given page	Learn whether the page has been indexed by Google; provides links for further URL information; this search can also alert you to possible site issues (duplicate content or possible DNS problems)	info: netconcepts.com will show you the page title and description, and invite you to view its related pages, incoming links, and page cached version
(Cache:)—See what the page looked like when Google crawled it	Shows Google's saved copy of the page	Google's <i>text</i> version of the page works the same way as SEO Browser	cache: www.stonetemple.com
(~keyword)	Shows keywords Google thinks are related to <i>keyword</i>	Can be very useful in uncovering related words that you should include on your page about <i>keyword</i>	~zoo ~trip will show you keywords related to <i>zoo</i> and <i>trip</i>

NOTE

When using the **site:** operator, some indexed URLs might not be displayed (even if you use the “repeat the search with omitted results included” link to see the full list). The **site:** query is notoriously inaccurate. You can obtain a more accurate count of the pages of your site indexed by Google by appending **&start=990&filter=0** to the URL of a Google result set for a **site:** command.

This tells Google to start with result 990, which is the last page Google will show you since it limits the results to 1,000. This must take place in two steps. First, enter a basic **site:yourdomain.com** command, and then get the results. Then go up to the address bar and append the **&start=990&filter=0** parameters to the end of the URL. Once done with this, you can look at the total pages returned to get a more accurate count.

To see more results, you can also use the following search patterns:

- [site:yourdomain.com/subdirectory1] + [site:yourdomain.com/subdirectory2] + etc. (the “deeper” you dig, the more/more accurate results you get)
- [site:yourdomain.com inurl:keyword1] + [site:yourdomain.com inurl:keyword2] + etc. (for subdirectory-specific keywords)
- [site:yourdomain.com intitle:keyword1] + [site:yourdomain.com intitle:keyword2] + etc. (for pages using the keywords in the page title)

Combined Google queries

To get more information from Google advanced search, it helps to learn how to effectively combine search operators. **Table 2-3** illustrates which search patterns you can apply to make the most of some important SEO research tasks.

TABLE 2-3. Combined Google search operators

What for	Description	Format	Example
Competitive analysis	Search who mentions your competitor; use the date range operator within Google advanced search to find the most recent brand mentions; the following brand-specific search terms can be used: [domainname.com], [domain name], [domainname], [site owner name], etc.	[domainname.com - site:domainname.com] (+ add &as_qdr=d [past one day] to the query string); use d3 for three days, m3 for three months, etc.	seomoz -site:seomoz.org during past 24 hours
Keyword research	Evaluate the given keyword competition (sites that apply proper SEO to target the term)	[inanchor: <i>keyword</i> intitle: <i>keyword</i>]	inanchor:seo intitle:seo

What for	Description	Format	Example
	Find more keyword phrases	[key * phrase]	free * tools
SEO site auditing	Learn whether the site has canonicalization problems	[site:domain.com - inurl:www]	site:iceposter.com -inurl:www
	Find the site's most powerful pages	[www site:domain.com]	www site:alchemistmedia.com
		[tld site:domain.tld]	org site:netconcepts.com
		[inurl:domain site:domain.com]	inurl:stonetemple site:stonetemple.com
		[domain site:domain.com]	alchemistmedia site:alchemistmedia.com
	Find the site's most powerful page related to the keyword	[site:domain.com keyword]	site:netconcept.com seo
		[site:domain.com intitle:keyword]	site:stonetemple.com intitle:seo
		[site:domain inanchor:keyword]	site:seomoz.org inanchor:seo
Link building	Find authority sites offering a backlink opportunity	[site:org bookmarks/links/"favorite sites/"] [site:gov bookmarks/links/"favorite sites/"] [site:edu bookmarks/links/"favorite sites/"]	site:org donors
	Search for relevant forums and discussion boards to participate in discussions and probably link back to your site	[inurl:forum OR inurl:forums keyword]	inurl:forum OR inurl:forums seo

Firefox plug-ins for quicker access to Google advanced search queries

You can use a number of plug-ins with Firefox to make accessing these advanced queries easier:

- [Advanced Dork](#), for quick access to (intitle:), (inurl:), (site:), and (ext:) operators for a highlighted word on a page, as shown in [Figure 2-24](#)
- [SearchStatus](#), for quick access to a (site:) operator to explore a currently active domain, as shown in [Figure 2-25](#)

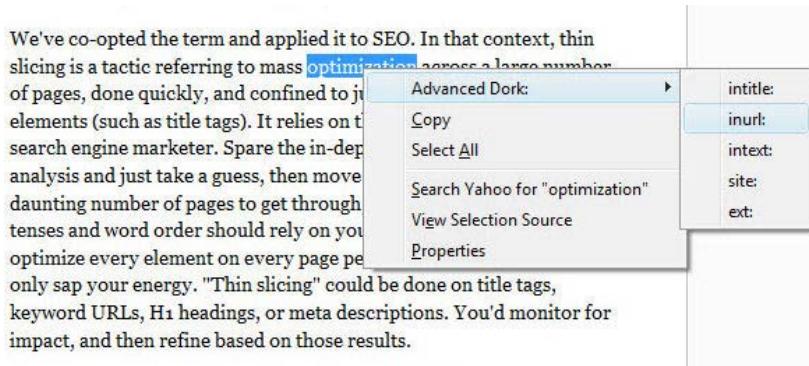


FIGURE 2-24. Advanced Dork plug-in for Firefox



FIGURE 2-25. SearchStatus plug-in for Firefox

Yahoo! Advanced Search Operators

The most popular of Yahoo!'s SEO [advanced search operators](#) is the one that checks the page or domain backlinks. Since Bing dropped its inbound link-checking ability except through its API, and while Google allows for only partial reports except for verified sites within Webmaster Central, Yahoo! has remained the main source for page backlink information.

Other Yahoo! queries are mostly used to enhance (link:) and linkdomain: searches. [Table 2-4](#) lists the most essential Yahoo! advanced operators (that differ from the ones detailed for Google) and samples of their usage in SEO. Note, however, that it is unclear which of these will be supported in the future due to the announced plan for Yahoo! search to be powered

by Microsoft's Bing (http://topics.nytimes.com/top/news/business/companies/yahoo_incl/yahoo-microsoft-deal/index.html).

TABLE 2-4. Yahoo! advanced search operators

Operator	Short description	SEO application	Examples
(Link:)—URL backlinks restricted search	Finds any <i>page</i> backlinks (if used in isolation, redirects to Yahoo! Site Explorer)	Find most the relevant backlinks of any page	link: http://www.seomoz.org/blog seo
(LinkDomain:)—Domain backlinks restricted search	Finds any <i>domain</i> backlinks (if used in isolation, redirects to Yahoo! Site Explorer)	Find the most relevant backlinks of any domain	linkdomain: www.alchemistmedia.com seo
(OriginUrlExtension:)—File type restricted search (Region:)—Area restricted search	Narrows search results to the pages that end in a particular file extension Narrows search results to the sites belonging to a particular territory	A few possible extensions/file types: PDF (Adobe Portable Document Format) HTML or <i>.htm</i> (Hypertext Markup Language) <i>.xls</i> (Microsoft Excel) <i>.ppt</i> (Microsoft PowerPoint) <i>.doc</i> (Microsoft Word)	originurlextension:pdf region:europe region:africa region:asia region:centralamerica region:downunder region:mediterranean region:mideast region:northamerica region:southamerica region:southeastasia

Operator	Short description	SEO application	Examples
(Url:)—URL restricted search	Finds a specific document in Yahoo's index	Learn whether the page was indexed by Yahoo!	url:http://www.netconcept.com

Combined Yahoo! Queries

Yahoo! also allows you to create combined search queries. [Table 2-5](#) shows some of the more unique ways to do this.

TABLE 2-5. Combined Yahoo! search operators

What for	Description	Format	Example
Keyword research	Explore territory-specific usage of your keyword	<i>keyword</i> region:europe	seo region:europe
SEO site auditing	Search for the best optimized page of the site	intitle: <i>keyword</i> inurl: <i>keyword</i> site:www.site.org	intitle:seo inurl:seo site:www.stonetemple.com
Link building	Find .edu or .gov sites linking to the page	link:http:// www.domain.com site:.edu OR site:.gov	link:http://www.amazon.com site:.edu OR site:.gov
	Find geospecific backlinks of a page/domain	link:http:// www.domain.com region:europe	link:http://www.amazon.com region:europe

Firefox extensions to help with Yahoo! (Link:) and (LinkDomain:) search operators

A couple of Firefox plug-ins help make it a bit easier to access these advanced Yahoo! search operators:

- [SearchStatus](#), to show currently active page and domain backlinks in Yahoo!, as shown in [Figure 2-26](#)
- [SEO for Firefox Toolbar Extension](#), which will show you detailed backlink information directly in your search results (see [Figure 2-27](#))
- [Seoquake Firefox Extension](#), which can be used to access Yahoo! backlink data (and a lot more) right from the search results page or from the currently active page (Seoquake), as shown in [Figure 2-28](#)



FIGURE 2-26. SearchStatus getting external backlinks from Yahoo!

[Running Shoes.com](http://www.runningshoes.com)
Running shoes and running apparel from Asics, Brooks, New Balance and Saucony
Women's Shoes - Men's Clearance - How to pick a shoe
www.runningshoes.com/ - Cached - Similar

#1 | PR: ? | Google Cache Date: ? | Traffic Value: ? | Age: ? | del.icio.us: ? | del.icio.us Page Bookmarks: ? | Digg: ? | Digg's Popular Stories: ? | Stumbleupon: ? | Twitter: ? | YI Links: ? | YI.edu Links: ? | YI.gov Links: ? | YI Page Links: ? | YI.edu Page Links: ? | Technorati: ? | Alexa: ? | Compete.com Rank: ? | Compete.com Unique: ? | Trends | Cached: ? | dmoz: ? | Bloglines: ? | Page blog links: ? | dir.yahoo.com: ? | Baidu: ? | Whois | SiteMap | Yahoo position: ? | Majestic SEO Linkdomain: ?

[Running Shoe Reviews: RunnersWorld.com](http://www.runnersworld.com/topic/07122_s6-240-400-0_0.html)
Need running shoes? Use the Runner's World Running Shoe Finder, read our extensive running shoe reviews, and find the best sneakers for you.
www.runnersworld.com/topic/07122_s6-240-400-0_0.html - Cached - Similar

#2 | PR: ? | Google Cache Date: ? | Traffic Value: ? | Age: ? | del.icio.us: ? | del.icio.us Page Bookmarks: ? | Digg: ? | Digg's Popular Stories: ? | Stumbleupon: ? | Twitter: ? | YI Links: ? | YI.edu Links: ? | YI.gov Links: ? | YI Page Links: ? | YI.edu Page Links: ? | Technorati: ? | Alexa: ? | Compete.com Rank: ? | Compete.com Unique: ? | Trends | Cached: ? | dmoz: ? | Bloglines: ? | Page blog links: ? | dir.yahoo.com: ? | Baidu: ? | Whois | SiteMap | Yahoo position: ? | Majestic SEO Linkdomain: ?

FIGURE 2-27. SEO for Firefox Toolbar Extension



FIGURE 2-28. Seoquake backlink data

Bing Advanced Search Operators

Although Bing is not as widely used for SEO purposes, it can still help you better understand your site's SEO issues or effectively find valuable SEO data. Besides, Bing offers several unique search operators worth looking into, as shown in [Table 2-6](#).

TABLE 2-6. Bing advanced operators

Operator	Short description	SEO application	Examples
(LinkFromDomain:)—Domain outbound links restricted search	Finds all pages the given domain links out to	Find most relevant sites your competitor links out to	linkfromdomain:seomoz.org seo
(Contains:)—File type restricted search	Narrows search results to pages linking to a document of the specified file type	Find pages linking to a specific document type containing relevant information	contains:wma seo
(IP:)—IP address restricted search	Shows sites sharing one IP address	ip:xxx.xxx.xxx.xxx	ip:207.182.138.245
(InBody:)—Body text keyword restricted search	Restricts the results to documents containing query word(s) in the body text of a page	Find pages containing the most relevant/best optimized body text	inbody:seo (equivalent to Google's intext:)
(Location:)/(loc:)—Location specific search	Narrows search results to a specified location (multiple location options can be found under Bing's advanced search)	Find geospecific documents using your keyword	seo loc:AU
(Feed:)—Feed keyword restricted search	Narrows search results to terms contained in RSS feeds	Find relevant feeds	feed:seo
(Hasfeed:)—Feed keyword restricted search	Narrows search results to pages linking to feeds that contain the specified keywords	Find pages linking to relevant feeds	hasfeed:seo site:cnn.com

More Advanced Search Operator Techniques

You can also use more advanced SEO techniques to extract more information.

Keyword difficulty

When building a web page, it can be useful to know how competitive the keyword is that you are going after, yet this can be difficult to obtain. The `intitle:` operator shows pages that are more focused on your search term than the pages returned without that operator (e.g., `intitle:"dress boots"`).

You can use different ratios to give you a sense of how competitive a keyword market is (higher results mean that it is more competitive). For example:

```
dress boots (20,900,000) versus "dress boots" (424,000) versus intitle:"dress
boots" (37,000)
Ratio: 20,900/37 = 565:1
Exact phrase ratio: 424/37 = 11:1
```

Another significant parameter you can look at is the `inanchor:` operator; for example, `inanchor:"dress boots"`. You can use this operator in the preceding equation instead of the `intitle:` operator.

Using number ranges

The number range operator can help restrict the results set to a set of model numbers, product numbers, price ranges, and so forth. For example:

```
site:stevespanglerscience.com "product/1700..1750"
```

Unfortunately, the number range combined with `inurl:` is not supported. So, the product number must be on the page. The number range operator is also great for copyright year searches (to find abandoned sites to acquire). Combine it with the `intext:` operator to improve the signal-to-noise ratio; for example, `intext:"copyright 1993..2005" -2008 blog`.

Advanced doc type search

The `filetype:` operator is useful for looking for needles in haystacks. Here are a couple of examples:

```
confidential business plan -template filetype:doc
forrester research grapevine filetype:pdf
```

Determine listing age

You can label results with dates that give a quick sense of how old (and thus trusted) each listing is; for example, by appending the `&as_qdr=m199` parameter to the end of a Google SERP URL, you can restrict results to those within the past 199 months.

Uncover subscriber-only or deleted content

You can get to subscriber-only or deleted content from the Cached link in the listing in the SERPs or by using the `cache:` operator. Don't want to leave a footprint? Add `&strip=1` to the end of the Google cached URL. Images on the page won't load.

If no Cached link is available, use Google Translate to take your English document and translate it from Spanish to English (this will reveal the content even though no Cached link is available):

```
http://translate.google.com/translate?prev=&hl=en&u=URL-GOES-HERE&sl=es&tl=en
```

Identify neighborhoods

The `related:` operator will look at the sites linking (the "Linking Sites") to the specified site, and then see which other sites are commonly linked to by the Linking Sites. This list is usually limited to between 25 and 31 results. These are commonly referred to as *neighborhoods*, as there is clearly a strong relationship between sites that share similar link graphs.

Find Creative Commons (CC) licensed content

Use the `as_rights` parameter in the URL to find Creative Commons licensed content. Here are some example scenarios to find CC-licensed material on the Web:

Permit commercial use

```
http://google.com/search?as_rights=(cc_publicdomain|cc_attribute|cc_sharealike|cc_nonderived).-(cc_noncommercial)&q=KEYWORDS
```

Permit derivative works

```
http://google.com/search?as_rights=(cc_publicdomain|cc_attribute|cc_sharealike|cc_noncommercial).-(cc_nonderived)&q=KEYWORDS
```

Permit commercial and derivative use

```
http://google.com/search?as_rights=(cc_publicdomain|cc_attribute|cc_sharealike).-(cc_noncommercial|cc_nonderived)&q=KEYWORDS
```

Make sure you replace `KEYWORDS` with the keywords that will help you find content that is of relevance to your site. The value of this to SEO is an indirect one. Creative Commons content can potentially be a good source of content for a website.

Vertical Search Engines

Vertical search is the term people sometimes use for specialty or niche search engines that focus on a limited data set (as already mentioned, Google calls them onebox results). Examples of vertical search solutions provided by the major search engines are image, video, news, and blog searches. These may be standard offerings from these vendors, but they are distinct from the engines' general web search functions.

Vertical search results can provide significant opportunities for the SEO practitioner. High placement in these vertical search results can equate to high placement in the web search results, often above the traditional 10 blue links presented by the search engines.

Vertical Search from the Major Search Engines

The big three search engines offer a wide variety of vertical search products. Here is a partial list:

Google

Google Maps, Google Images, Google Product Search, Google Blog Search, Google Video, Google News, Google Custom Search Engine, Google Book Search, Google US Gov't Search, etc.

Yahoo!

Yahoo! News, Yahoo! Local, Yahoo! Images, Yahoo! Video, Yahoo! Shopping, Yahoo! Audio Search, etc.

Bing

Bing Image, Bing Video, Bing News, Bing Maps, Bing Health, Bing Products, etc.

Image search

All three search engines offer image search capability. Basically, image search engines limit the data that they crawl, search, and return in results to images. This means files that are in GIF, TIF, JPG, and other similar formats. [Figure 2-29](#) shows the image search engine from Bing.

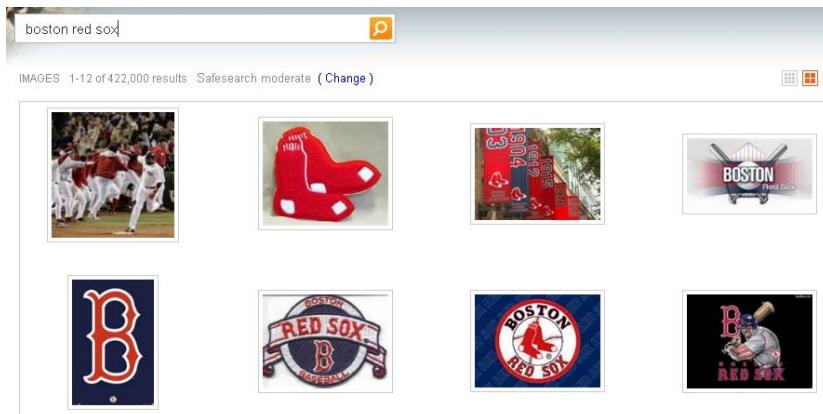


FIGURE 2-29. Image search results from Bing

Image search engines get a surprisingly large number of searches performed on them. According to comScore, more than 1 billion image searches were performed in October 2008, or a little more than 8.3% of all searches performed in that month. Similar data from Nielsen

Online shows image search comprised 6.0% of all search in January 2009. However, since an image is a binary file, it cannot be readily interpreted by a search engine crawler.

The search engine has to rely on text surrounding the image, the alt attribute within the img tag, and the image filename. Optimizing for image search is its own science, and we will discuss it in more detail in “[Optimizing for Image Search](#)” on page 346 in Chapter 8.

Video search

As with image search, video search engines focus on searching specific types of files on the Web, in this case video files, such as MPEG, AVI, and others. [Figure 2-30](#) shows a quick peek at video search results from YouTube.

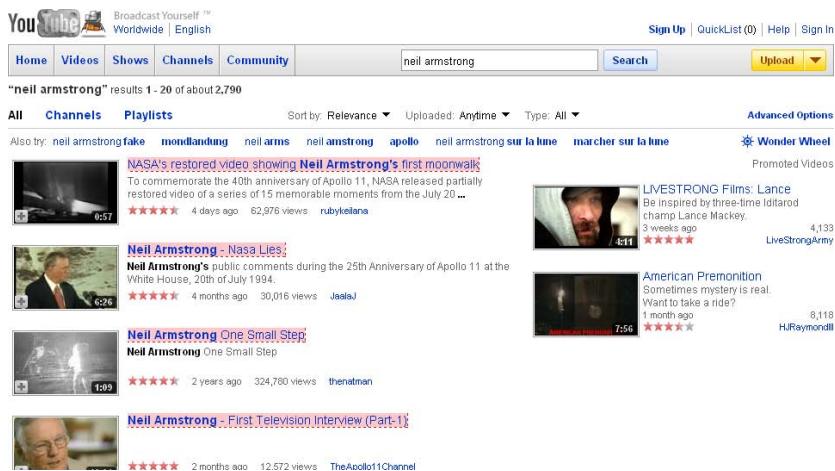


FIGURE 2-30. Video search results from YouTube

A very large number of searches are also performed in video search engines. Hitwise and comScore data shows approximately 125 million searches performed on video search on the major search engine properties (e.g., <http://video.google.com>, <http://video.yahoo.com>, and <http://video.bing.com>) in October 2008, and then this number balloons to 2.6 billion searches once you include YouTube (<http://www.youtube.com>), which has become the #2 search engine on the Web.

There is significant traffic to be gained by optimizing for video search engines and participating in them. Once again, these are binary files and the search engine cannot easily tell what is inside them.

This means optimization is constrained to data in the header of the video and on the surrounding web page. We will discuss video search optimization in more detail in “[Others: Mobile, Video/Multimedia Search](#)” on page 366 in Chapter 8.

However, each search engine is investing in technology to analyze images and videos to extract as much information as possible. For example, OCR technology is being used to look for text within images, and other advanced technologies are being used to analyze video content. Flesh-tone analysis is also in use to detect porn or recognize facial features. The application of these technologies is in its infancy, and is likely to evolve rapidly over time.

News search

News search is also unique. News search results operate on a different time schedule. News search results have to be very, very timely. Few people want to read the baseball scores from a week ago when several other games have been played since then.

News search engines must be able to retrieve information in real time and provide near instantaneous responses. Modern consumers tend to want their news information now.

Figure 2-31 is a quick look at the results from a visit to Yahoo! News.

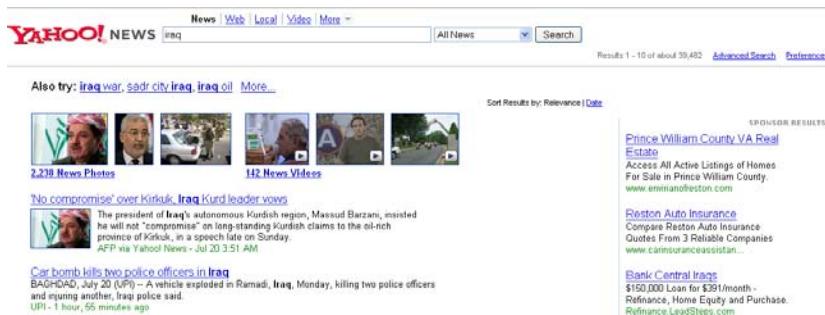


FIGURE 2-31. News search results from Yahoo!

As with the other major verticals, there is a lot of search volume here as well. To have a chance of receiving this volume, you will need to become a news source. This means timely, topical news stories generated on a regular basis. There are other requirements as well, and we will discuss them further in “[Optimizing for News, Blog, and Feed Search](#)” on page 355 in [Chapter 8](#).

Local search/maps

Next up in our hit parade of major search verticals is local search (a.k.a. map search). Local search engines search through databases of locally oriented information, such as the name, phone number, and location of local businesses around the world, or just provide a service, such as offering directions from one location to another. **Figure 2-32** shows Google Maps local search results.

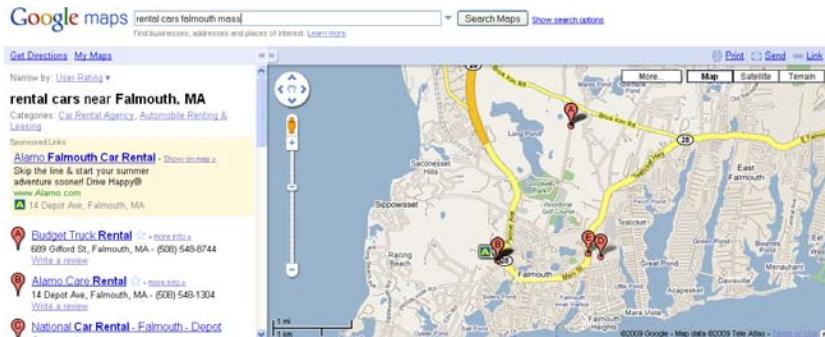


FIGURE 2-32. Local search results from Google

The integration of local search results into regular web search results has dramatically increased the potential traffic that can be obtained through local search. We will cover local search optimization in detail in “Optimizing for Local Search” on page 338 in Chapter 8.

Blog search

Google has implemented a search engine focused just on blog search called Google Blog Search (misnamed because it is an RSS feed engine and not a blog engine). This search engine will respond to queries, but only search blogs (more accurately, feeds) to determine the results. Figure 2-33 is an example search result for the search phrase *barack obama*.

The screenshot shows the Google Blog Search interface for the query "barack obama". At the top, there is a search bar with the query, followed by three buttons: "Search Blogs", "Search the Web", and "Advanced Blog Search Preferences". Below the search bar, the results are categorized under "Blog results". The results are displayed in a list format, with each item showing the title, author, publication time, and a brief description. There are also sections for "Related Blogs" and "Blog Top Stories".

- Blog results**
- Browse Top Stories**
- Published**
 - Last hour
 - Last 12 hours
 - Last day
 - Past week
 - Past month
 - Anytime
 - [Choose Dates](#)
- Related Blogs:**
 - [Barack Obama and Joe Biden: The Change We Need](#) - http://www.barackobama.com/
 - [Change.gov - The Official Web Site of the](#) - http://change.gov/
 - [Flickr: Barack Obama's Photostream](#) - http://www.flickr.com/photos/barackobamadotcom/?PHPSESSID=2bae7...
 - [Barack Obama - Pictures, Videos, Breaking News](#) - http://www.huffingtonpost.com/news/barack-obama/
 - [The White House, President Barack Obama](#) - http://www.whitehouse.gov/
- Barack Obama's window: 'Prime-time' The Swamp**
2 hours ago by Mark Silva
Six months in office today, President **Barack Obama** hopes to start drawing on a reservoir of public support for his own leadership as well as public frustration over the state of health care in America to press Congress for sweeping ...
[The Swamp](#) - http://www.swamppolitics.com/news/politics/blog/
- Barack Obama Twitter Addiction » Right Pundits**
6 hours ago by MCCAIN
A **Barack Obama** Twitter account? Surely you are joking! But worse, we at Right Pundits are afraid to report that our president is more infatuated with mind-dumbing technology than almost anyone else in the video-game era ...
[Right Pundits](#) - http://www.rightpundits.com/ - References
[More results from Right Pundits]

FIGURE 2-33. Results from Google Blog Search

We explore the subject of optimizing for Google Blog Search in “[Optimizing for News, Blog, and Feed Search](#)” on page 355 in [Chapter 8](#).

Book search

The major search engines also offer a number of specialized offerings. One highly vertical search engine is Google Book Search, which specifically searches only content found within books, as shown in [Figure 2-34](#).



FIGURE 2-34. Google Book Search

Job search

Yahoo! also has a number of vertical search products. Yahoo! hotjobs is an example of a product designed to allow people to search for jobs (see [Figure 2-35](#)).

Celebrity xRank

Microsoft also has some unique vertical search properties. One of the more interesting ones is Celebrity xRank, which offers data on celebrity rankings and search trends, as shown in [Figure 2-36](#).

Universal Search/Blended Search

Google made a big splash in 2007 when it announced Universal Search. This was the notion of integrating images, videos, and results from other vertical search properties directly into the main web search results.

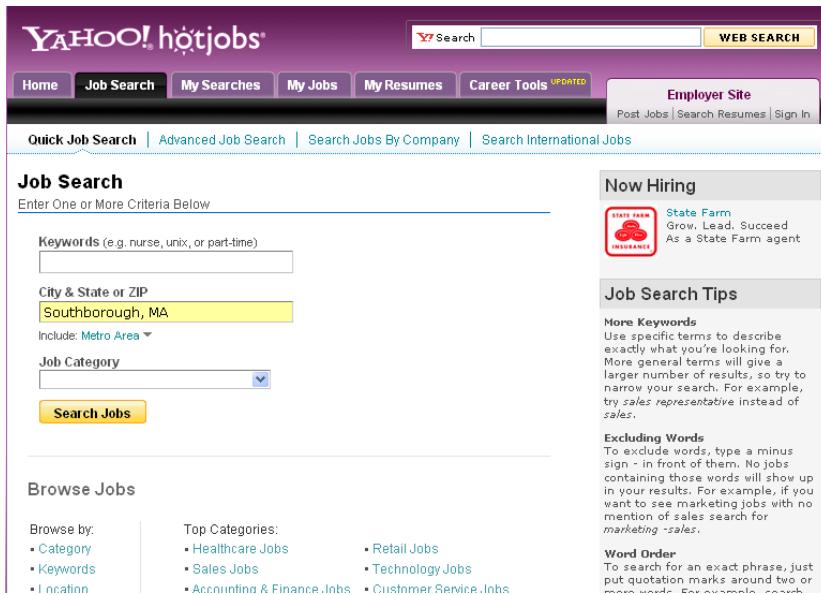


FIGURE 2-35. Yahoo! hotjobs

This quickly became a huge driver of traffic to sites because of their images, videos, news, local search information, and more. The other search engines quickly followed suit and began offering vertical search integration before 2007 was over. People now refer to this general concept as Blended Search (since Universal Search is specifically associated with Google). A look at some Universal Search results from Google can help illustrate the concept (see Figure 2-37).

Note the news results, along with an image at the very top of the results, along with more image results farther down. This information is coming from Google's news search index. If you look farther down in the search results, you will continue to see more vertical results, including video results and a timeline (see Figure 2-38).

A wide range of vertical data sets have been integrated into Google's Universal Search, as well as into the Blended Search results of the other search engines. In addition to the preceding examples, you can also see images, videos, and local data integrated into the traditional web search results.

The advent of Blended Search has significantly increased the opportunity for publishers with matching vertical data sets (such as a rich music library) to gain significant additional traffic to their sites by optimizing these data sets for the appropriate vertical search.



FIGURE 2-36. Bing xRank search

Meta search

Meta search engines are search engines that aggregate results from multiple search engines and present them to the user. The two most well-known ones are [MetaCrawler](#) and [Dogpile](#). However, their cumulative search volume is quite small, and these do not factor into SEO strategies.

More specialized vertical search engines

Vertical search can also come from third parties. Here are some examples:

- Comparison shopping engines, such as PriceGrabber, Shopzilla, and NexTag
- Travel search engines, such as Expedia, Travelocity, Kayak, and Uptake
- Real estate search engines, such as Trulia and Zillow
- People search engines, such as Spock and Wink
- Job search engines, such as Indeed, CareerBuilder, and SimplyHired
- Music search engines, such as iTunes Music Store

- B2B search engines, such as [Business.com](#), KnowledgeStorm, Kellysearch, and ThomasNet

The screenshot shows a Google search results page for the query "neil armstrong". At the top, there's a search bar with the query, a "Search" button, and links for "Advanced Search" and "Preferences". Below the search bar, a "Web" section header is followed by a "Show options..." link. The search results are categorized under "News results for neil armstrong", "Text results for neil armstrong", "Image results for neil armstrong", and "More results from en.wikipedia.org".

News results for neil armstrong

- Armstrong** moon line was mine: scientist - 14 hours ago
A British scientist claims he came up with astronaut **Neil Armstrong**'s famous moon walking quote while working at Australia's deep-space tracking station. ...
[The Age](#) - 5159 related articles »
- Former Tyler Mayor Layton Recalls Meeting **Neil Armstrong** - Tyler Morning Telegraph
Charting nasa's future: to the moon and beyond - 89.3 KPCC - 265 related articles »

Neil Armstrong - Wikipedia, the free encyclopedia

Neil Alden Armstrong (born August 5, 1930) is an American aviator and a former astronaut, test pilot, university professor, and United States Naval Aviator. ...
[en.wikipedia.org/wiki/Neil_Armstrong](#) - Cached - Similar

Apollo 11 - Wikipedia, the free encyclopedia

Comparison between the original and the newly restored Apollo 11 video showing **Neil Armstrong** making his way to the lunar surface, by climbing down the ...
[en.wikipedia.org/wiki/Apollo_11](#) - Cached - Similar
[More results from en.wikipedia.org](#) »

Neil Armstrong

Biography of the test pilot who's first space flight occurred in 1966 aboard Gemini 8.
[starchild.gsfc.nasa.gov/docs/StarChild/.../armstrong.html](#) - Cached - Similar

Image results for neil armstrong - Report images

Four small square images showing Neil Armstrong in his white space suit, likely from different media sources.

FIGURE 2-37. Google Universal Search results

In addition, some companies offer products that allow anyone to build his own search engine, such as [Google's Custom Search Engines](#), [Eurekster](#), and [Rollyo](#). Also, specialty search engines are offered by the major search engines not covered in this section.

There is an enormous array of different vertical search offerings from the major search engines, and from other companies as well. It is to be expected that this explosion of different vertical search properties will continue.

Effective search functionality on the Web is riddled with complexity and challenging problems. Being able to constrain the data types (to a specific type of file, a specific area of interest, a specific geography, or whatever) can significantly improve the quality of the results for users.

Video results for **neil armstrong**

First Moon Landing 1969
1 min 44 sec
www.youtube.com



Neil Armstrong, 60 Minutes
Interview
14 min
video.google.com

Astronaut Bio: **Neil Armstrong**

Neil Armstrong (NASA Photo) **Neil A. Armstrong** NASA Astronaut (former). PERSONAL DATA: Born August 5, 1930 in Wapakoneta, Ohio. Married. Two sons. ...
www.jsc.nasa.gov/Bios/htmlbios/armstrong-na.html - [Cached](#) - [Similar](#)

NASA - **Armstrong, Neil**

Born in 1930, **Neil A. Armstrong**, a United States astronaut, was the first person to set foot on the moon. On July 20, 1969, **Armstrong** and Buzz Aldrin landed ...
www.nasa.gov/worldbook/armstrong_neil_worldbook.html - [Cached](#) - [Similar](#)

NASA - Biography of **Neil Armstrong**

Neil A. Armstrong, the first man to walk on the moon, was born in Wapakoneta, Ohio, on August 5, 1930. He began his NASA career in Ohio. ...
www.nasa.gov/centers/glenn/about/bios/neilabio.html - [Cached](#) - [Similar](#)

Neil Armstrong Biography - Biography of **Neil Armstrong** - First Man ...

Born on August 5, 1930 in Auglaize County, OH, **Neil Armstrong** was the eldest of 3 children. **Neil Armstrong** worked at numerous jobs around town and at the ...
space.about.com/od/astronautbiographies/.../neilarmstrong.htm - [Cached](#) - [Similar](#)

Never-before seen photo shows **Neil Armstrong's** face as he first ...

Jul 22, 2009 ... An amazing new photograph showing **Neil Armstrong's** face through his space suit visor has come to light.
www.dailymail.co.uk/.../Never-seen-photo-shows-Neil-Armstrongs-face-walks-moon.html - 57 minutes ago - [Similar](#)

Timeline results for **neil armstrong**

- 1969 On **July 20, 1969**, Apollo 11 astronauts **Neil Armstrong** and Edwin "Buzz" Aldrin climbed inside their lunar lander, undocked from the command module ...
books.google.com
- 1969 On **July 20, 1969**, **Neil Armstrong** became the first human being to set foot on the moon. This book tells the story of **Armstrong's** life, with details ...
www.rif.org

FIGURE 2-38. More Google Universal Search results

Country-Specific Search Engines

At this stage, search is truly global in its reach. Google is the dominant search engine in many countries, but not all of them. How you optimize your website depends heavily on the target market for that site, and the search engine(s) that is (are) the most important in that market.

According to comScore, Google is receiving 62% of all searches performed worldwide. In addition, Google is the market share leader in every regional market. Most notable, though, is the Asia Pacific region, where Google holds a relatively narrow 36.1% to 26.8% edge over Baidu, the largest search engine in China. This is the only regional market in which Google

has less than 60% market share, and it also happens to be the largest market for search in the world (in terms of total searches performed).

Here is some data on countries where other search engines are the major players:

China

Baidu News reported in February 2009 that Baidu had more than 73% market share in China in 2008 (<http://baidu-news.blogspot.com/2009/02/baidus-market-share-at-732-domestic.html>).

Russia

According to figures announced by Yandex, the company's market share in Russia comprises about 50% of all searches (<http://www.searchengineworld.com/seo/3459183.htm>).

South Korea

Naver was estimated by the *International Herald Tribune* (<http://www.iht.com/articles/2007/07/04/business/naver.php>) to have about 77% market share in South Korea in mid-2007.

Czech Republic

The StartupMeme Technology blog reported that **Seznam** was reported to have more than 60% market share in Czechoslovakia in mid-2008 (<http://startuppeme.com/google-to-acquire-czech-search-engine-seznam-for-1-billion/>).

Optimizing for Specific Countries

One of the problems international businesses continuously need to address with search engines is identifying themselves as "local" in the eyes of the search engines. In other words, if a search engine user is located in France and wants to see where the wine shops are in Lyons, how does the search engine know which results to show?

Here are a few of the top factors that contribute to international ranking success:

- Owning the proper domain extension (e.g., .com.au, .co.uk, .fr, .de, .nl) for the country that your business is targeting
- Hosting your website in the country you are targeting (with a country-specific IP address)
- Registering with local search engines:
 - Google: <http://www.google.com/local/add/>
 - Yahoo!: <http://listings.local.yahoo.com/csubmit/index.php>
 - Bing: <https://ssl.bing.com/listings/BusinessSearch.aspx>
- Having other sites from the same country link to you
- Using the native language on the site (an absolute requirement for usability)
- Placing your relevant local address data on every page of the site
- Defining your preferred region in Google Webmaster Tools

All of these factors act as strong signals to the search engines regarding the country you are targeting, and will make them more likely to show your site for relevant local results.

The complexity increases when targeting multiple countries. We will discuss this in more depth in “[Best Practices for Multilanguage/Country Targeting](#)” on page 275 in [Chapter 6](#).

Profile of China’s Internet Usage

Based on the data that China released in July 2008, China has more Internet users than the United States, with 253 million users. Here is a summary of data from the [Chinese report](#):

- China has the most Internet users in the world (see [Figure 2-39](#)).

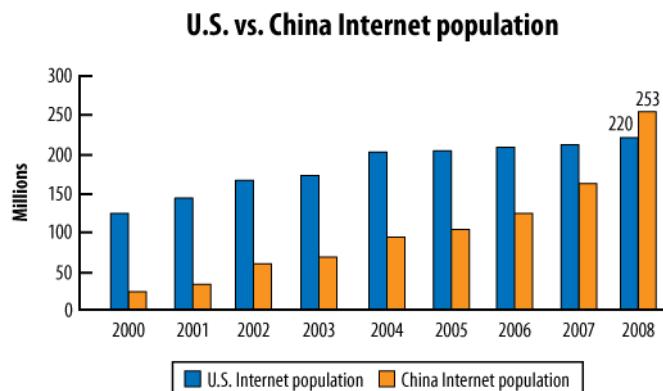


FIGURE 2-39. U.S. versus China Internet users

- China has the most broadband users in the world (see [Figure 2-40](#)). *“This report, the 22nd Statistical Report on the Internet Development in China, also indicates the number of broadband users has reached 214 million, which also tops the world.”*
- China’s Internet penetration rate continues to grow and grow and grow (see [Figure 2-41](#)):
 - U.S. Internet usage has hovered around a 70% penetration rate in the past five years, while Chinese Internet penetration has jumped from 7% to almost 20% in the same period.
 - China could plausibly reach a similar penetration rate to the United States within 20 years.
- Mobile phone use in China continues to grow:
 - China has *601 million* mobile phone users, according to the [latest government report](#).
 - From January 2008 to June 2008, there were *53.3 million* new mobile phone users.

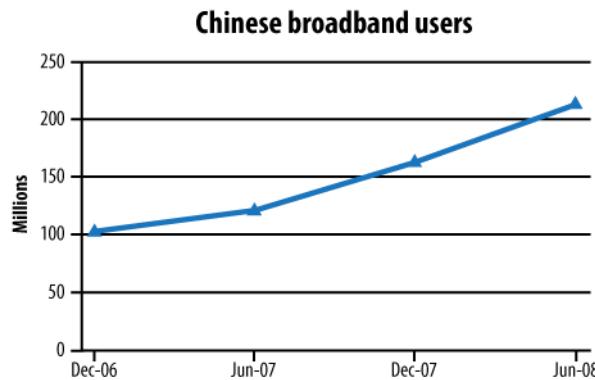


FIGURE 2-40. Chinese broadband users

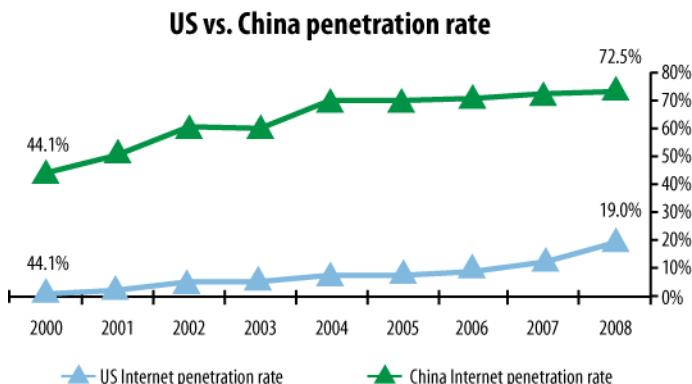


FIGURE 2-41. U.S. versus China Internet penetration

- One carrier, China Mobile, has more than 414 million mobile subscribers and is ranked #1 in the world (http://en.wikipedia.org/wiki/List_of_mobile_network_operators).
- Only 12% of these users have accessed the Internet from a mobile phone because of the lack of a proper 3G network (none of the Chinese telcos have a 3G license); an estimated 73 million have accessed the Internet from a mobile phone.
- However, the Chinese government is moving aggressively to issue 3G licenses to the major telcos.

Conclusion

Understanding how search engines work is an important component of SEO. The search engines are constantly tuning their algorithms. For that reason, the successful SEO professional is constantly studying search engine behavior and learning how they work.

CHAPTER THREE

Determining Your SEO Objectives and Defining Your Site's Audience

Setting SEO Goals and Objectives

SEO, ONCE A HIGHLY SPECIALIZED TASK RELEGATED TO THE BACK ROOMS of a website development team, is now a mainstream marketing activity. This dramatic rise can be attributed to three emerging trends:

- Search engines drive dramatic quantities of focused traffic, comprising people intent on accomplishing their research and purchasing goals. Businesses can earn significant revenues by leveraging the quality and relevance of this traffic for direct sales, customer acquisition, and branding/awareness campaigns.
- Visibility in search engines creates an implied endorsement effect, where searchers associate quality, relevance, and trustworthiness with sites that rank highly for their queries.
- Dramatic growth in the interaction between offline and online marketing necessitates investment by organizations of all kinds in a successful search strategy. Consumers are increasingly turning to the Web before making purchases in verticals such as real estate, autos, furniture, and technology. Organizations cannot afford to ignore their customers' needs as expressed through searches conducted on Google, Yahoo!, and Bing.

Search engine optimization is a marketing function, and it needs to be treated like one. SEO practitioners need to understand the services, products, overall business strategy, competitive landscape, branding, future site development, and related business components just as much as members of other marketing divisions, whether online or offline.

Like any other marketing function, it is important to set specific goals and objectives—and if the goal is not measurable, it is not useful. Setting up such objectives is the only way you can determine whether you are getting your money's worth from your SEO effort. And although SEO can be viewed as a *project*, the best investment, in our opinion, is to treat it as more of a *process*—one which is iterative, is ongoing, and requires steady commitment from the stakeholders of an organization.

Viewing SEO like PPC (like something you decidedly turn on and off) is like viewing a healthy diet as something you do only when you are overweight, as opposed to eating a healthy diet as a lifestyle choice. Too heavy? Crash diet. PPC too expensive? Pause the campaigns. The tactic may work in the right application, but with SEO, those with the most success are those who view site optimization as a lifestyle choice. The results may not appear instantly, but they will handsomely reward a business after patient and prudent commitment.

Strategic Goals SEO Practitioners Can Fulfill

Although SEO is not a cure-all for businesses, it can fit into a company's overall business strategy in several critical ways.

Visibility (branding)

Consumers assume that top placement in the search engines is like a stamp of approval on a business. These users believe that surely the company could not rank highly in the search engines if the company were not one of the best in its field, right?

If you are an experienced search engine user, you probably recognize that the preceding statement is not true. However, the fact is that many consumers, and even business searchers, sometimes interpret search results as an implicit endorsement.

Therefore, for critical brand terms, the SEO should work toward improving the search engine rankings for the website she is working on. There is a subtlety here, though. Few businesses will need help for their company name; that is, if your company name is Acme Widget Co., you will most likely rank #1 for that search term even with little SEO effort. There are a few reasons for this, one of the most important being that many of your inbound links to your site will use your company name as the anchor text, and very few links will be given to other websites using your company name as the anchor text.

However, if you sell solar panels, you will want to rank well for the search term *solar panels*. When users see you ranking highly on that search term, they will assume you are one of the best places to buy solar panels.

SEO for branding is about ranking highly for the generic search terms that relate to the purpose of your website.

Website traffic

Long gone are the days of a “build it and they will come” paradigm on the Web. Today’s environment is highly competitive, and you need great SEO to ensure targeted, high-quality traffic to your site.

Of course, a business that engages with many of its customers through offline channels can tell them to visit their website to drive traffic. But the SEO practitioner fills the different, more critical role of bringing new prospects to your website from an audience of people who would not otherwise have been interested in, or perhaps aware of, the business at all.

Experienced SEO practitioners know that users search for products, services, and information using an extraordinarily wide variety of search queries and query types. An SEO professional performs keyword research (which we will discuss in [Chapter 5](#)) to determine which search queries people actually use. For example, when searching for a set of golf clubs, it may be that some users will type in *lefthanded golf clubs* as a search query.

The person who enters *lefthanded golf clubs* as a search query may not even know that such a company exists until she performs that search. Or if she does know that one exists, it was apparently not top of mind enough for her to seek the company’s website out directly.

Capturing that traffic would provide the company with incremental sales of its golf clubs that it probably would not have gotten otherwise. Knowing that, the SEO process works on a site architecture strategy (see [Chapter 6](#)) and a link-building strategy (we cover this in [Chapter 7](#)) to help the site’s pages achieve competitive search engine rankings for these types of terms.

High ROI

Branding and traffic are nice, but the most important goal is to achieve the goals of your organization. For most organizations, that means sales, leads, or advertising revenue. For others, it may mean the promotion of a particular message. An important component of SEO is to deliver not just traffic, but relevant traffic that has the possibility of converting. The great thing about SEO is that it can result in dramatically improved website ROI. Whether you are selling products and services, advertising and looking for branding value, or trying to promote a specific viewpoint to the world, a well-designed SEO strategy can result in a very high return on investment when contrasted with other methods of marketing.

For many organizations, SEO brings a higher ROI when compared to TV, print, and radio. Traditional media is not in danger of being replaced by SEO, but SEO can provide some high-margin returns that complement and enhance the use of offline media. Data released by SEMPO in early 2009 shows that organic SEO is considered one of the very highest ROI activities for businesses (see [Figure 3-1](#)).

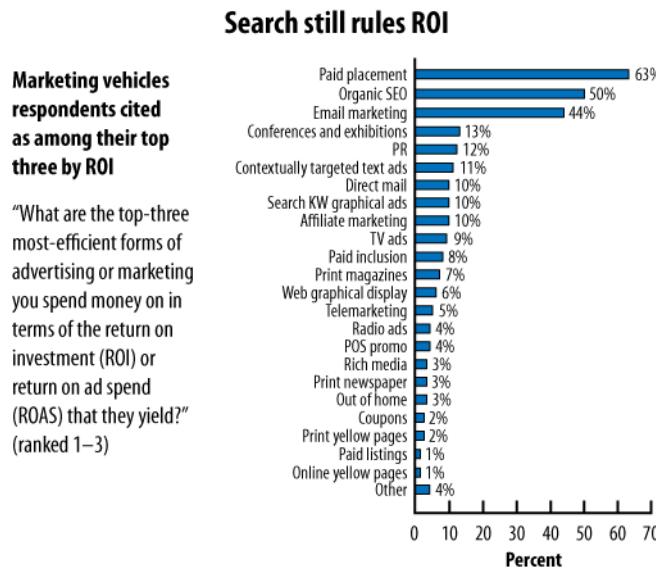


FIGURE 3-1. SEO, a high ROI activity

In addition, a growing number of businesses operate purely online. Two examples of these are [Amazon](#) and [Zappos](#).

Every SEO Plan Is Custom

There is no such thing as a cookie-cutter SEO plan, and for this, all parties on the SEO bandwagon should rejoice. The ever-changing, dynamic nature of the search marketing industry requires constant diligence. SEO professionals must maintain a research process for analyzing how the search landscape is changing, because search engines strive to continuously evolve to improve their services and monetization. This environment provides search engine marketers a niche within which currency and demand for their services are all but guaranteed for an indefinite period of time; and it provides advertisers the continuous opportunity, either independently or through outside consulting, to achieve top rankings for competitive target searches for their business.

Organizations should take many factors into account when pursuing an SEO strategy, including:

- What the organization is trying to promote
- Target market
- Brand
- Website structure

- Current site content
- Ease with which the content and site structure can be modified
- Any immediately available content
- Available resources for developing new content
- Competitive landscape
- And so on...

Learning the space the business is in is not sufficient. You may have two businesses offering the same products on the market, but it may not make sense for them to use the same SEO strategy.

For example, if one of the two competitors put its website up four years ago and the other company is just rolling one out now, the second company may need to focus on specific vertical areas where the first company's website offering is weak.

The first company may have an enormous library of written content that the second company would struggle to replicate and extend, but perhaps the second company is in a position to launch a new killer tool that the market will like.

Do not underestimate the importance of your SEO plan. The only thing you can do by skipping over this process or not treating it seriously is to short-sell the business results for your company.

Understanding Search Engine Traffic and Visitor Intent

As we discussed in “[The Mission of Search Engines](#)” on page 2 in [Chapter 1](#), searchers enter many different types of searches. These are typically classified into three major categories of activity:

Navigational query

This is a query with the intent to arrive at a specific website or page (e.g., the person types in your company name, *Acme Device Co.*).

Informational query

This is a search performed to receive an answer to a broad or direct question with no specific source in mind (e.g., *Celtics game score*).

Transactional query

A person who types in *digital camera* may be looking to buy one now, but it is more likely that she is researching digital cameras. This is an example of an initial transactional query, which can evolve in stages. For example, here are some other types of transactional queries that occur at a later stage in the buying cycle:

- The user types in *buy digital camera*. Although there is no information in the query about which one she wants to buy, the intent still seems quite clear.
- The searcher types in *canon powershot G10*. The chances are very high that this user is looking to buy that particular camera.

The geographic location of the searcher can also be very valuable information. For example, you may want to show something different to a searcher in Seattle than to a searcher in Boston.

Part of an SEO plan is to understand how the various relevant types of searches relate to the content and architecture of your website.

Developing an SEO Plan Prior to Site Development

It is widely understood in the industry that search engine optimization should be built in, as early as possible, to the entire site development strategy, from choosing a content management system (CMS) and planning site architecture to on-page content development. As you will see in [Chapter 6](#), SEO practitioners have significant input in both of these areas. Of course, many businesses learn about the need for SEO only after they have built the site, in which case the time to start is now.

SEO plans have many moving parts, and SEO decisions can have a significant impact on other departments such as development, other marketing groups, and sales. Getting that input as soon as possible will bring the best results for a business at the least possible cost (imagine that you develop your site and learn you need to replace the CMS—that would be very, very painful!).

Business Factors That Affect the SEO Plan

Here are some examples of business issues that can impact SEO:

Revenue/business model

It makes a difference to the SEO practitioner if the purpose of the site is to sell products, sell advertising, or obtain leads. We will discuss this more in the later sections of this chapter.

Target customers

Who are you trying to reach? This could be an age group, a gender group, or as specific as people looking to buy a house within a 25-mile radius of Orlando.

Competitor strategies

The competitive landscape is another big factor in your SEO plan. Competition may be strongly entrenched in one portion of the market online, and it may make sense to focus on a different segment. Or you may be the big dog in your market but you have specific competitors you want to fend off.

Branding goals

There may be terms that are critical for you to own for branding reasons.

Budget for content development

An important part of link building (which we will discuss in detail in [Chapter 7](#)) is the quality of your content, as well as your capacity to commit to the ongoing development of quality on-page site content.

How your potential customers search for products like yours

Understanding what customers do when they are searching for products or services like yours is one of the most basic functions of SEO (we will discuss it in detail in [Chapter 5](#)). This involves mapping the actual search queries your target customers use when they go to a search engine to solve their current problem.

Understanding Your Audience and Finding Your Niche

A nontrivial part of an SEO plan is figuring out who you are targeting with your website. This is not always that easy to determine. As you will see in this section, many factors enter into this, including the competition, the particular strengths or weaknesses of your own company, and more.

Mapping Your Products and Services

Successful SEO requires a thorough understanding of the business organization itself. What products, services, and types of information and resources does your organization have to offer?

As we outlined in the preceding section, a critical SEO activity is to understand who is searching for what you are trying to promote, which requires thoroughly understanding all aspects of your offering. You will also need to understand the broad market categories that your products fall into, as each of these categories might relate to sections of your website that you may want to create. By having sections of the site for those categories, you create an opportunity to obtain search traffic related to those categories.

You also should consider business development and the company's expansion strategy at the outset of the SEO planning process. Consider Amazon, which began as a bookseller but has evolved into a general purpose e-tailer. Sites that go through these types of changes may need to be substantially restructured, and such restructurings can be a source of major SEO headaches. Anticipating those changes in advance provides the opportunity to recommend architectural approaches to dealing with those changes.

Content Is King

One aspect of determining the desired audience for your website is determining who you want to reach, which requires an understanding of what you have to offer visitors to your site, both now and in the future.

You may have a deep library of “how to” content, great videos, a unique photo gallery, or an awesome tool that people are interested in using. Each of these can be valuable in building a world-class website that does well in the search engines.

The content you have available to you will affect your keyword research and site architecture, as your site content is the major source of information that search engines use to determine what your site is about. As we discussed in “[Algorithm-Based Ranking Systems: Crawling, Indexing, and Ranking](#)” on page 30 in Chapter 2, you need relevant content to even be “in the game” in search (i.e., if someone searches for *lefthanded golf clubs* and you don’t have any content related to lefthanded golf clubs, chances are good that you won’t rank for that search query).

As we will discuss in [Chapter 7](#), on-site content also affects your link-building efforts. Link building is very similar to PR in that the success of your link-building efforts is integrally related to what you are promoting (i.e., what are you asking them to link to?).

Consider Site A, a site that has built a really solid set of articles on a given topic. However, 20 other sites out there have an equally solid set of articles on the same topic, and many of these other sites have been in the major search engine indexes for much longer than Site A.

Site A has a serious problem. Why would someone link to it? There is nothing new there. Chances are that Site A will succeed in getting some links to its articles; however, it will likely never be able to establish itself as a leader because it has nothing new to offer.

To establish itself as a leader, Site A must bring something new and unique to the market. Perhaps it can offer a solution to a problem that no one else has been able to solve before. Or perhaps it covers the same content as its competition, but it is the first to release a high-quality video series on the topic. Or perhaps it focuses on a specific vertical niche, and establishes itself as a leader in that specific niche.

One of the most important decisions Site A’s leadership needs to make is where and how they are going to establish themselves as one of the top experts and resources in their market space. If they plan to make their website a major player in capturing market-related search engine traffic, this is not an optional step.

When looking at content plans it is critical to consider not only what you already have, but also what you could develop. This relates to budget for resources to build the content. A publisher with no budget to spend on content development has few choices that she can make in her SEO plan, whereas another publisher who has a team of in-house content developers looking for something to do has a lot more options.

As a result, a critical part of the SEO planning process is to map the SEO and business goals of the website to the available budget to add new content, and to prioritize the list of opportunities to estimate the size of the ROI potential.

Segmenting Your Site's Audience

Let's not forget the audience itself! This is very important background information for the SEO practitioner. For example, Site A may be a website that sells gadgets. As a result, the site's developers go out and implement a brilliant campaign to rank for the terms they consider relevant. Being young and energetic, they focus on the way their peers search for gadgets, but Site A is focused on selling gadgets to people who are age 50 or older.

Uh-oh, Site A is in trouble again. Why? One reason it may be in trouble is that the target audience for Site A (the over-50 crowd) may use different search terms to search for gadgets than the younger generation does, and now Site A is bringing in search traffic from people who are not interested in its products, and not bringing in traffic from those who might be!

Similar things can happen with gender. For example, women and men may not search for their shoes the same way, as shown in [Figure 3-2](#), which lists the top shoe-related search terms from [Wordtracker](#).

womens shoes	<input type="checkbox"/> mens shoes
womens wide shoes	<input type="checkbox"/> mens dress shoes
womens dress shoes	<input type="checkbox"/> mens golf shoes
womens wide width shoes	<input type="checkbox"/> dolce and gabbana mens shoes
womens running shoes	<input type="checkbox"/> dr scholls mens shoes
large size womens shoes	<input type="checkbox"/> mens casual shoes
buy womens designer shoes online	<input type="checkbox"/> mens shoes wide widths
womens golf shoes	<input type="checkbox"/> get new balance mens shoes
womens designer shoes	<input type="checkbox"/> gucci mens shoes
extra wide womens shoes	<input type="checkbox"/> mens dress shoes in chicago
womens dc shoes	<input type="checkbox"/> mens water shoes
womens high heel shoes	<input type="checkbox"/> mens gucci shoes
discount womens shoes	<input type="checkbox"/> mens work shoes
womens size 12 shoes	<input type="checkbox"/> mens sperry shoes

FIGURE 3-2. Difference in search by men versus women

As you can see in [Figure 3-2](#), search terms used can vary significantly by gender.

Another major criterion to consider might be location. Searchers in Austin, Texas may naturally want a different version of your product than searchers in Chicago. For that matter,

because they want different products, they may use different search terms, which requires extensive keyword research—yet another critical aspect of the SEO process.

Advanced Methods for Planning and Evaluation

There are many methodologies for business planning. One of the more well-known ones is the SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis. There are also methodologies for ensuring that the plan objectives are the right type of objectives, such as the SMART (Specific, Measurable, Achievable, Realistic, Timelined) plan. Next we will take a look at both of these in the context of SEO.

SWOT analysis

Sometimes you need to get back to the basics and carry out a simple overview strategy of where you are in the marketplace, and where you would like to be. A simple SWOT analysis is a great starting point. It creates a grid from which to work and is very simple to execute.

As you can see from the SWOT chart in [Figure 3-3](#), Strengths and Weaknesses usually stem from internal (on-site, business operational, business resource) sources, whereas Opportunities and Threats are from external sources.

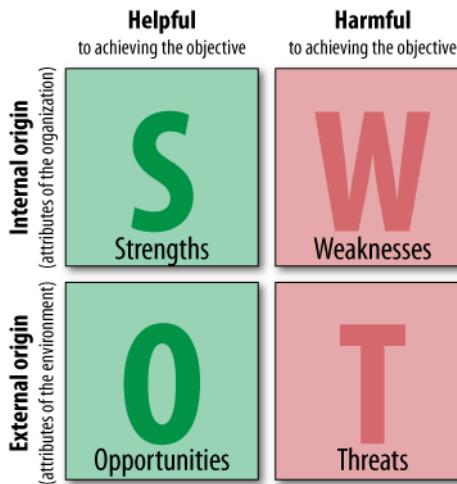


FIGURE 3-3. Example SWOT chart

Where does SEO fit in here? To explore this, it is helpful to use an example. Take Business X. It has a website that was built on WordPress, makes use of category tagging, adds at least one page of content every two days, and has excellent knowledge of its industry. Its domain name isn't ideal—Businessnameandkeyword.com—but it is decent.

Business X does not get much traffic from search engines, but its rival, Business Y, does because Business Y has had its website up for a long period of time and received some great links along the way. Business Y doesn't have any SEO plan and relies on its main page to bring in all traffic. This is because Business Y has a keyword-rich domain name and people have linked to the site using the domain name (giving it keyword-rich anchor text), and because of its longevity on the Web.

There aren't a lot of target search queries; in fact, there are fewer than 50,000 for the core set of keywords. Business X's site ranks on the second page of Google results, whereas Business Y is ranked #3, with Wikipedia and [About.com](#) taking up the top two positions.

Neither of the businesses is spending money on PPC (paid search) traffic, and the niche doesn't have much room for other entrants (there may be 10 to 15 competitors). Both sites have similar link authority in terms of strengths and numbers. The businesses deal in impulse purchases—the products evoke strong emotions.

[Figure 3-4](#) shows what the SWOT for Business X might look like.

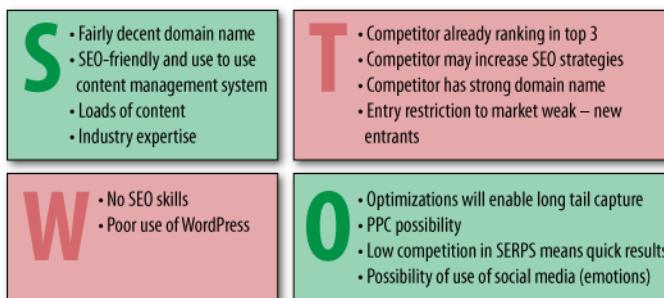


FIGURE 3-4. Sample SWOT chart data for Business X

The preceding analysis suggests quick wins for the Business X site, as well as where the priorities are. It also forms a great starting point for a long-term strategy and tactical maneuvers. This example is simplistic, but it illustrates how instructive a fleshed out SWOT can be. It does require you to have analyzed your site, the main competitor(s), the keywords, and the search engine results pages (SERPs).

Get SMART

Every company is unique, so naturally their challenges are unique. Even a second SEO initiative within the same company is not the same as the first initiative. Your initial SEO efforts have changed things, creating new benchmarks, new expectations, and different objectives. These all make each SEO project a new endeavor.

One way to start a new project is to set SMART objectives. Let's look at how to go about doing that in the world of SEO.

Specific objectives are important. It is easy to get caught up in details of the plan and lose sight of the broader site objectives. You may think you want to rank #1 for this phrase or that, but in reality you want more customers. Perhaps you don't even need more customers, but you want higher sales volume, so in fact having the same number of orders but with a higher average order value would meet your objectives better.

Measurable objectives are essential if one is to manage the performance in meeting them—you can't manage what you can't measure. SEO practitioners have to help their clients or organizations come to grips with analytics, and not just the analytics software, but the actual processes of how to gather the data, how to sort it, and most important, how to use it to make informed decisions.

Achievable objectives are ones which can be accomplished with the available resources. You could decide to put a man on Mars next year, for example, but it is just too big an undertaking to be feasible. You can be ambitious, but it is important to pick goals that can be met. You cannot possibly sell to more people than exist in your market. There are limits to markets, and at a certain point the only growth can come from opening new markets, or developing new products for the existing market.

Aside from basic business achievability, there are also limits to what can rank at #1 for a given search query. The search engines want the #1 result to be the one that offers the most value for users, and unless you are close to having the website that offers the most value to users, it may be unreasonable to expect to get to that position, or to maintain it if you succeed in getting there.

Realistic objectives are about context and resources. It may be perfectly achievable to meet a certain objective, but only with greater resources than may be presently available. Even a top ranking on the most competitive terms around is achievable for a relevant product, but it is realistic only if the resources required for such an effort are available.

Time-bound is the final part of a SMART objective. If there is no timeline, no project can ever fail, since it can't run out of time. SEO generally tends to take longer to implement and gather momentum than paid advertising. It is important that milestones and deadlines be set so that expectations can be managed and course corrections made.

"We want to rank at #1 for loans" is not a SMART objective. It doesn't identify the specific reason why the company thinks a #1 ranking will help it. It doesn't have a timeline, so there is no way to fail. It doesn't state an engine on which to be #1, so there's a guaranteed argument if it means to rank on MSN and you get the job done on Yahoo!.

"To increase approved loan applications generated by natural search by 30% over six months" is a far better objective. There is a deadline, and the company can certainly gauge progress toward the specific objective. The company can look at its current market share and the resources committed to see whether this is achievable and realistic.

SEO for Raw Traffic

Optimizing a site for search engines and creating keyword-targeted content helps a site rank for key search terms, which typically leads to direct traffic and referring links as more and more people find, use, and enjoy what you've produced. Thousands of sites on the Web leverage this traffic to serve advertising, directly monetizing the traffic sent from the engines. From banner ads to contextual services such as Google's AdSense program to affiliate programs and beyond, web advertising has become a massive industry—\$25 billion plus, according to eMarketer (<http://www.thestandard.com/predictions/u-s-online-advertising-grows-10-2009>).

Here are some factors to think about when considering SEO for raw traffic:

When to employ

Use it when you can monetize traffic without actions or financial transactions on your site (usually through advertising).

Keyword targeting

Keyword targeting in this scenario can be very broad. The goal here isn't typically to select specific keywords, but rather to create lots of high-quality content that naturally targets interesting/searched-for terms. Instead of singular optimization on specific terms, the focus is on accessibility and best practices throughout the site to earn traffic through both high volume and long tail queries (for more on what *long tail* is, see [Chapter 5](#)).

Concentrate efforts on great content, and use keyword-based optimization only as a secondary method to confirm the titles/headlines of the works you create.

Page and content creation/optimization

A shallow, highly crawlable link structure is critical to getting all of your content indexed—follow good information architecture practices (for more on this, see “[Creating an Optimal Information Architecture](#)” on page 187 in [Chapter 6](#)) and use intelligent, detailed category and subcategory structures to get the most benefit out of your work. You'll also need to employ good on-page optimization (titles, headlines, internal linking, etc.) and make your articles easy to share and optimized for viral spreading (see “[Root Domains, Subdomains, and Microsites](#)” on page 201 and “[Optimization of Domain Names/URLs](#)” on page 207 in [Chapter 6](#) for more on this).

SEO for E-Commerce Sales

One of the most direct monetization strategies for SEO is driving relevant traffic to an e-commerce shop to boost sales. Search traffic is among the best quality available on the Web, primarily because a search user has expressed a specific goal through her query, and when this matches a product or brand the web store carries, conversion rates are often extremely high.

Forrester Research estimated the e-commerce market would top \$235 billion in 2009 (http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9061108&taxonomyId=14&intsrc=kc_top), though the recent economic downturn may affect that number

somewhat. With so many dollars flowing over the Web, it is little surprise that e-commerce-focused SEO is among the most competitive and popular applications of the practice.

Here are some factors to think about when considering SEO for e-commerce sales:

When to employ

Use it when you have products/services that are directly for sale on your website.

Keyword targeting

Paid search advertising is an excellent way to test the efficacy and potential ROI of keyword targets. Find those that have reasonable traffic and convert well, and then pursue them further. You'll often find that the more specific the query—brand-inclusive, product-inclusive, and so on—the more likely the visitors are to make the purchase.

Page and content creation/optimization

You'll typically need some serious link building, along with internal optimization, to achieve high rankings for competitive, high-value keywords that bring in conversion-focused traffic. Manual link building is an option here, but scalable strategies that leverage a community or customers can be equally, or even more, valuable.

SEO for Mindshare/Branding

A less popular but equally powerful application of SEO is its use for branding purposes.

Bloggers, social media/community websites, content producers, news outlets, and dozens of other web publishing archetypes have found tremendous value in appearing atop search results and using the resulting exposure to bolster their brand recognition and authority.

The process is fairly simple, much like the goals in traditional advertising of ad repetition to enter the buyer's consideration set. (Read about the three laws of branding at <http://www.palgrave-journals.com/bm/journal/v16/n3/full/2550139a.html> for more information on this topic.) Similarly, online marketers have observed that being at the top of search rankings around a particular subject has a positive impact on traffic, consideration, and perceived authority.

Here are some factors to think about when considering SEO for mindshare/branding:

When to employ

Use it when branding, or communicating a message, is your goal. If you do not have direct monetization goals for the moment or for the foreseeable future, this is the approach for you.

Keyword targeting

A keyword focus is less critical here—you'll likely have a few broad terms that receive the high traffic you want, but the long tail may be far more achievable and the better target. Focus on keywords that are going to bring you visitors who are likely to be interested in and remember your brand.

Page and content creation/optimization

Make an accessible site, use good link structure, apply best practices, and focus on links for domain authority rather than chasing after specific keywords.

SEO for Lead Generation and Direct Marketing

Although lead generation via the Web is less direct than an e-commerce transaction, it is arguably just as valuable and important for building customers, revenue, and long-term value. Millions of search queries have commercial intents that can't be (or currently aren't) fulfilled directly online. These can include searches for services such as legal consulting, contract construction, commercial loan requests, alternative energy providers, or virtually any service or product people source via the Web.

Here are some factors to think about when considering SEO for lead generation and direct marketing:

When to employ

Use it when you have a non-e-commerce product/service/goal that you want users to accomplish on your site or for which you are hoping to attract inquiries/direct contact over the Web.

Keyword targeting

As with e-commerce, choose phrases that convert well, have reasonable traffic, and have previously performed in PPC campaigns.

Page and content creation/optimization

Although you might think it would be easier to rank high in the SERPs for lead generation programs than for e-commerce, it is often equally challenging. You'll need a solid combination of on-site optimization and external link building to many different pages on the site (with good anchor text) to be competitive in the more challenging arenas.

SEO for Reputation Management

Since one's own name—whether personal or corporate—is one's identity, establishing and maintaining the reputation associated with that identity is generally of great interest.

Imagine that you search for your brand name in a search engine and high up in the search results is a web page that is highly critical of your organization. Consider the search for *Whole Foods* shown in [Figure 3-5](#).

SEO for reputation management is a process for neutralizing negative mentions of your name in the SERPs. In this type of SEO project, you would strive to occupy additional spots in the top 10 results to push the critical listing lower and hopefully off the first page. You may accomplish this using social media, major media, bloggers, your own sites and subdomains, and various other tactics.

Whole Foods Market: Natural and Organic Grocery

Jul 19, 2009 ... **Whole Foods** Market is the world's largest retailer of natural and organic foods, with stores throughout North America and the United Kingdom.

[+ Show stock quote for WFM](#)

www.wholefoodsmarket.com/ - Cached - Similar -   

[Stores](#)

[Health & Nutrition](#)

[Careers](#)

[Catering](#)

[Products](#)

[Company](#)

[Recipes](#)

[Customer Service](#)

[More results from wholefoodsmarket.com »](#)

Whole Foods Market Stores | WholeFoodsMarket.com

Whole Foods Market teamed up with musician and activist Sheryl Crow to create a special edition of our signature reusable shopping bag to help raise ...

www.wholefoodsmarket.com/stores/ - Cached - Similar -   

Whole Foods Market (**WholeFoods**) on Twitter

Fresh organic tweets from **Whole Foods** Market HQ in Austin, TX.
twitter.com/WholeFoods - Cached - Similar -   

Local business results for **whole foods** near Massachusetts - Change location



-  **Whole Foods** Market - www.wholefoodsmarket.com
647 Washington St, Newton, MA - (617) 965-2070
 8 reviews, directions, hours, and more »
 -  **Whole Foods** - www.wholefoodsmarket.com
115 Prospect St, Cambridge, MA - (617) 492-0070
5 reviews, directions, hours, and more »
 -  **Whole Foods** Market - www.wholefoodsmarket.com
916 Walnut St, Newton, MA - (617) 969-1141
5 reviews, directions, and more »
-  [More results near Massachusetts »](#)

Whole Foods Market - Wikipedia, the free encyclopedia

Whole Foods Market (NASDAQ: WFM) is an Austin, Texas-based foods grocer. As of September 2008, the company operates over 275 locations in the United States ...
en.wikipedia.org/wiki/Whole_Foods_Market - Cached - Similar -   

Whole foods - Wikipedia, the free encyclopedia

Mar 12, 2009 ... **Whole foods** are those that are unprocessed and unrefined, or processed and refined as little as possible before being consumed. ...
en.wikipedia.org/wiki/Whole_foods - Cached - Similar -   

What smells at **Whole Foods**? | SocialistWorker.org

May 7, 2009 ... **Whole Foods**' corporate leaders claim to operate according to a.
socialistworker.org/2009/05/07/smells-at-whole-foods - Cached - Similar -   

WFM: Summary for **Whole Foods** Market, Inc. - Yahoo! Finance

Get detailed information on **Whole Foods** Market, Inc. (WFM) including quote performance, Real-Time ECN, technical chart analysis, key stats, ...
finance.yahoo.com/q?s=WFM - Cached - Similar -   

FIGURE 3-5. Reputation management search example

SEO enables this process through both content creation and promotion via link building. Although reputation management is among the most challenging of SEO tasks (primarily because you are optimizing many results for a query rather than one), demand for these types of services is rising as more and more companies become aware of the issue.

Here are some factors to think about when considering SEO for reputation management:

When to employ

If you're trying to either protect your brand from negative results appearing on page 1 or push down already existing negative content, reputation management SEO is the only path to success.

Keyword targeting

Chances are this is very easy—it's your personal name, your brand name, or some common variant (and you already know what it is). You might want to use keyword research tools just to see whether there are popular variants you're missing.

Page and content creation/optimization

Unlike the other SEO tactics, reputation management involves optimizing pages on many different domains to demote negative listings. This involves using social media profiles, public relations, press releases, and links from networks of sites you might own or control, along with classic optimization of internal links and on-page elements. It is certainly among the most challenging of SEO practices, especially in Google, where query deserves diversity (QDD) can mean you have to work much harder to push down negatives because of how the algorithm favors diverse content.

SEO for Ideological Influence

For those seeking to sway public (or private) opinion about a particular topic, SEO can be a powerful tool. By promoting ideas and content within the search results for queries likely to be made by those seeking information about a topic, you can influence the perception of even very large groups. Politicians and political groups and individuals are the most likely employers of this tactic, but it can certainly be applied to any subject from the theological to the technical or civic.

Here are some factors to think about when considering SEO for ideological influence:

When to employ

Use it when you need to change minds or influence decisions/thinking around a subject—for example, a group of theoretical physicists attempting to get more of their peers considering the possibility of alternative universes as a dark matter source.

Keyword targeting

It's tough to say for certain, but if you're engaging in these types of campaigns, you probably know the primary keywords you're chasing and can use keyword research query expansion to find others.

Page and content creation/optimization

This is classic SEO, but with a twist. Since you're engaging in ideological warfare in the SERPs, chances are you've got allies you can rally to the cause. Leverage your combined links and content to espouse your philosophical preferences.

Conclusion

To bring this all together, your objectives, tactics, and strategies need to be aligned. They need to take into account your market, your business, and the competition. The best strategy is the one that gets you to your goals the fastest. Don't spread yourself too thin. Remember to ask yourself the tough questions, such as:

- Does your company need direct sales, traffic, branding, or some combination of these?
- Are there specific influencers you're trying to reach with a message?
- Is the organization/brand subject to potentially negative material that needs to be controlled/mitigated?
- Do you have products/services you sell, either directly over the Web or through leads established online?

Getting such answers won't be easy, but it will be worth the effort.

CHAPTER FOUR

First Stages of SEO

SEO PROJECTS REQUIRE FORETHOUGHT AND PLANNING TO OBTAIN THE BEST RESULTS, and SEO needs to be considered during, and incorporated into, all stages of a website development or redevelopment project. For example, the site architecture (including the selection of a content management system or CMS), the marketing plan (including branding concepts), and much more are impacted.

In this chapter, we will discuss several aspects of how SEO projects start:

- Putting together an SEO plan
- Performing a technical SEO audit of a site
- Setting a baseline for measuring results and progress

These are the things you want to do at the very beginning of your SEO efforts for any website.

The Major Elements of Planning

As any experienced SEO consultant will tell you, you should incorporate your SEO strategy into the site planning process long before your site goes live. Your strategy should be well outlined before you make even the most basic technology choices, such as the hosting platform and your CMS. However, this is not always possible—and in fact, more often than not an SEO professional will be brought in to work on a site that already exists.

Regardless of when you start, there are a number of major components to any SEO plan that you need to address long before you research the first title tag.

Technology Choices

As we already suggested, SEO impacts major technology choices. For example, a CMS can facilitate—or, possibly, eliminate—your SEO strategy. Some platforms do not even allow you to have titles and meta descriptions that vary from one web page to the next, create hundreds (or thousands) of pages of duplicate content, or make a 302 (temporary) redirect the default redirect you need to use. All of these things could be disastrous for your website.

This problem also exists with web servers. For example, if you use IIS, the default redirect choice is a 302 (as we will explain in “[Redirects](#)” on page 190 in [Chapter 6](#), a 301 [permanent] redirect is essential for most redirect applications). You can configure IIS to use a 301 redirect, but this is something you need to understand and build into your SEO plan up front.

Market Segmentation

Another critical factor is the nature of the market in which you are competing. This tells you how competitive the environment is in general, and when you augment it with additional research, you can use this information to tell how competitive the SEO environment is.

In some markets, natural search is intensively competitive. For instance, [Figure 4-1](#) shows the Google results for *credit cards*. In this market, Visa, Master Card, American Express, and Discover all fail to make the #1 position in Google’s results, so you know this market is highly competitive.

This does not mean you should give up on the market, especially if it is already your business; however, you might choose to focus your SEO on less competitive terms that will still bring you many qualified leads.

Where You Can Find Great Links

As you will see in [Chapter 7](#), getting third parties to link their websites to yours is a critical part of SEO; without inbound links, there is little to no chance of ranking for competitive terms in search engines such as Google, whose algorithm relies heavily on link measuring and weighting criteria.

An early part of the SEO brainstorming process is identifying the great places to get links, as well as the types of content you might want to develop to encourage linking from other quality websites. Note that we, the authors, advocate pursuing few, relevant, higher-quality links over hundreds of low-quality links, as 10 good links can go much further than thousands of links from random blog posts or forums. Understanding this will help you build your overall content plan.

Credit Cards - Compare Credit Card Offers at CreditCards.com
 Jul 22, 2009 ... Compare Credit Cards & Credit Card Offers at CreditCards.com. Search credit cards and reviews about the best low interest, ...
www.creditcards.com/ - 3 hours ago - Cached - Similar -

Balance Transfer Cards	Airline Credit Cards
Bad Credit	Fair Credit
Instant Approval Cards	Limited or No Credit History
Low Interest Credit Cards	Excellent Credit

[More results from creditcards.com >](#)

Visa USA | More People Go with Visa
 Visa USA has information on Visa credit and debit cards, including Visa Classic, Visa Secured, Visa Gold, Visa Platinum, Visa Signature, and Visa Check ...
 Show stock quote for V
www.visa.com/ - Cached - Similar -

Capital One Credit Cards, Banking, Auto Loans, Checking and ...
 Information on Capital One credit cards, branch banking products, auto loans, checking, savings, and money market accounts and online banking. Show stock quote for COF
www.capitalone.com/ - Cached - Similar -

News results for credit cards

 Credit Card Disputes Tossed Into Disarray - 15 hours ago
 By ROBIN SIDEL and AMOL SHARMA Two major arbitration firms are backing away from the business of resolving disputes between customers and their credit-card ...
 Wall Street Journal - 189 related articles >
 Report: No banks easing credit card standards; 68% tightening - CreditCards.com - 82 related articles >
 Most Big Credit Card Firms Raising Rates, Fees - NPR - 96 related articles >

Discover Credit Cards : Rewards Credit Card, Travel Card, Student Card
 Browse for low APR Discover Credit Card including Discover More, Open Road Card, Student Credit Card, and Miles Card.
www.discovercard.com/ - Cached - Similar -

Welcome to MasterCard Worldwide
 MasterCard Worldwide manages a family of well-known, widely accepted payment cards brands including MasterCard, Maestro and Cirrus and serves financial ...
 Show stock quote for MA
www.mastercard.com/ - Cached - Similar -

Credit Cards - Citi® Cards
 Find Citi® Credit Cards that provide industry-leading benefits, services and rewards. Citi Credit Cards are tailored to fit any lifestyle.
www.citicards.com/ - Cached - Similar -

Credit Cards | Best Credit Card Offers | Low APR Interest Rewards ...
 Bankrate.com offers credit card offers and applications for the best low APR secured and unsecured travel reward credit cards.
www.bankrate.com/credit-cards.aspx - Cached - Similar -

American Express Credit Cards, Business Services, and Travel Services
 American Express offers individuals online access to its world-class Card, Financial, and Travel services, including financial advice, retirement planning, ...
<https://home.americanexpress.com/> - Cached - Similar -

CHASE® Credit Cards - Credit Card Offers, Travel Rewards Credit ...
 Compare Credit Cards & Apply Online- Find Low APR Cards, Gas Discount, Balance Transfer, Reward, Cash Back, Travel & Small Business Credit Cards.
<https://www.chase.com/.../card.../credit-cards-at-chase.html> - Cached - Similar -

An overview of all Bank of America credit cards by category for ...
 Compare credit cards, check preapproval offers and link to additional credit card activities and rewards information.
<https://www.bankofamerica.com/credit-cards/> - Cached - Similar -

FIGURE 4-1. Sample results for a competitive query

Content Resources

The driver of any heavy-duty link campaign is the quality and volume of your content. If your content is of average quality and covers the same information dozens of other sites have

covered, it will not attract many links. If, however, you are putting out quality content, or you have a novel tool that many will want to use, you are more likely to receive external links.

At the beginning of any SEO campaign you should look at the content on the site and the available resources for developing new content. You can then match this up with your target keywords and your link-building plans to provide the best results.

Branding Considerations

Of course, most companies have branding concerns as well. However, there may be strategies that you cannot pursue because of the nature of your brand, such as the following:

- Developing and promoting interesting articles on [Digg](#) is a strategy that many companies employ. Articles that are popular enough to make the front page of Digg get a very large number of visitors and a good number of links as a result. But the audience is skewed heavily toward 13- to 28-year-old males, and the type of content that is required to become popular with that audience might not fit with some brands (take, for example, the AARP).
- Some companies put up pages about other companies' products, most likely as part of a competitive comparison, to rank the other companies' brand name in the search engine. Again, this might be pushing the envelope a bit for some brands.

The list of situations where the brand can limit the strategy is quite long, and the opposite can happen too, where the nature of the brand makes a particular SEO strategy pretty compelling.

Competition

Your SEO strategy can also be influenced by your competitors' strategies, so understanding what they are doing is a critical part of the process for both SEO and business intelligence objectives. There are several scenarios you might encounter:

- The competitor discovers a unique, highly converting set of keywords.
- The competitor discovers a targeted, high-value link.
- The competitor saturates a market segment, justifying your focus elsewhere.
- Weaknesses appear in the competitor's strategy, which provide opportunities for exploitation.

Understanding the strengths and weaknesses of your competition from an SEO perspective is a significant part of devising your own SEO strategy.

Identifying the Site Development Process and Players

Before you start working on any SEO project, you need to know three things: who your target audience is, what your message is, and how your message is relevant. There are no web design

tools or programming languages that tell you these things. Marketing, advertising, and PR have to set the objectives before you can implement them—successful SEO requires a team effort.

Your SEO team should be cross-functional and multidisciplinary, consisting of the team manager, the technical team, the creative team, and the major stakeholders from marketing, advertising, and PR. In a smaller organization, you may have to wear all of those hats yourself.

The team leader wants to know who the target audience is. What does marketing know about them? How did we find them? What metrics will we use to track them? All of this is key information that should have an impact on the project’s technical implementation.

Advertising needs to have its messages well thought out and prepared. You do not want your team bickering over whether to optimize for “hardcore widget analysis” or “take your widgets to the next level.” Advertising serves multiple purposes, but its most fundamental purpose is to compel people to take a specific action: what action are you compelling people to take?

PR has to take your story to the media and entice them into writing and talking about it. What message do they want to deliver? You have to mirror that message in your content. If they say you’re relevant to white bears but your project plan says you’re relevant to African bees, the whole project is in trouble. When you’re creating visibility, the people who build up your brand have to see a clear, concise focus in what you do. If you provide them less than that, they’ll find someone else to talk about.

The technical and creative team is responsible for delivering the project. They take direction from marketing, advertising, and PR on what needs to be accomplished, but from there on out they have to put the pieces into place. As the project unfolds, marketing has to come back and say whether the target audience is being reached. Advertising has to come back and say the message is clear. PR has to come back and say the media likes what they see.

Ongoing feedback is essential because the success of your project is determined solely by whether you’re meeting your goals. A successful SEO team understands all of these interactions and is comfortable relying on each team member to do his part. Establishing good communications among team members is essential.

And even if you are a team of one, you need to understand all of these steps. Getting all aspects of the marketing problem addressed (as relates to SEO) is a requirement for success.

Defining Your Site’s Information Architecture

Whether you have built a site already or not, you should plan to research the desired site architecture (from an SEO perspective) at the start of your SEO project, a task which can be divided into two major components: technology decisions and structural decisions.

Technology Decisions

As we outlined previously in this chapter, your technology choices can have a major impact on your SEO results. The following is an outline of the most important issues to address at the outset:

Dynamic URLs

Although Google now states that dynamic URLs are not a problem for the company, this is not entirely true, nor is it the case for the other search engines. Make sure your CMS does not end up rendering your pages on URLs with many convoluted parameters in them.

Session IDs or user IDs in the URL

It used to be very common for your CMS to track individual users surfing your site by adding a tracking code to the end of the URL. Although this worked well for this purpose, it was not good for search engines, because they saw each URL as a different page rather than variants of the same page. Make sure your CMS does not ever serve up session IDs.

Superfluous flags in the URL

Related to the preceding two items is the notion of extra junk being present on the URL. This probably does not bother Google, but it may bother the other search engines, and it interferes with the user experience for your site.

Links or content based in JavaScript, Java, or Flash

Search engines often cannot see links and content implemented using these technologies. Make sure the plan is to expose your links and content in simple HTML text.

Content behind forms (including pull-down lists)

Making content accessible only after completing a form (such as a login) or making selections from improperly implemented pull-down lists is a great way to hide content from the search engines. So, do not use these techniques unless you want to hide your content!

Temporary (302) redirects

This is also a common problem in web server platforms or CMSs. The 302 redirect blocks a search engine from recognizing that you have intended to move content, and is very problematic for SEO. You need to make sure the default redirect your systems use is a 301, or understand how to configure it so that it becomes the default.

All of these are examples of basic technology choices that can adversely affect your chances for a successful SEO project. Do not be fooled into thinking that SEO issues are understood, let alone addressed, by all CMS vendors out there—unbelievably, many are still very far behind the SEO curve.

Also do not assume that all web developers understand the SEO implications of what they develop. Learning about SEO is not a requirement to get a software engineering degree or become a web developer (in fact, almost no known college courses address SEO). It is up to

you, the SEO expert, to educate the other team members on this issue as early as possible in the process.

Structural Decisions

One of the most basic decisions to make about a website concerns internal linking and navigational structures. What pages are linked to from the home page? What pages are used as top-level categories that then lead you to other related pages? Do pages that are relevant to each other link to each other? There are many, many aspects to determining a linking structure for a site, and it is a major usability issue because visitors make use of the links to surf around your website. For search engines, the navigation structure helps their crawlers determine what pages you consider the most important on your site, and it helps them establish the relevance of the pages on your site to specific topics.

[Chapter 6](#) covers site architecture and structure in detail. This section will simply reference a number of key factors that you need to consider before launching into developing or modifying a website.

Target keywords

As we will discuss in [Chapter 5](#), keyword research is a critical component of SEO. What search terms do people use when searching for products or services similar to yours? How do those terms match up with your site hierarchy? Ultimately, the logical structure of your pages should match up with the way users think about products and services like yours. [Figure 4-2](#) shows how this is done on the Amazon site.

Cross-link relevant content

Linking between articles that cover related material can be very powerful. It helps the search engine ascertain with greater confidence how relevant a web page is to a particular topic. This can be extremely difficult to do well if you have a massive e-commerce site, but Amazon solves the problem very well, as shown in [Figure 4-3](#).

The “Frequently Bought Together” and “What Do Customers Ultimately Buy After Viewing This Item?” sections are brilliant ways to group products into categories that establish the relevance of the page to certain topic areas, as well as create links between relevant pages.

In the Amazon system, all of this is rendered on the page dynamically, so it requires little day-to-day effort on Amazon’s part. The “Customers Who Bought...” data is part of Amazon’s internal databases, and the “Tags Customers Associate...” data is provided directly by the users themselves.

Of course, your site may be quite different, but the lesson is the same. You want to plan on having a site architecture that will allow you to cross-link related items.



FIGURE 4-2. Example of a well-thought-out site hierarchy

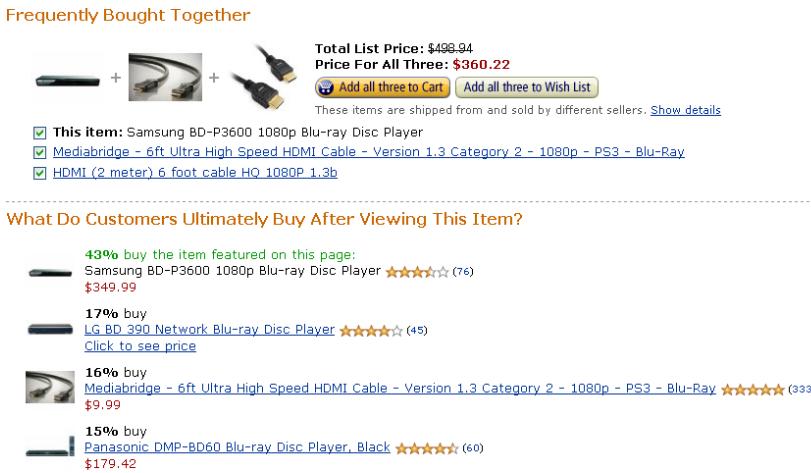


FIGURE 4-3. Product cross-linking on Amazon

Use anchor text

Anchor text is one of the golden opportunities of internal linking. As an SEO practitioner, you need to have in your plan from the very beginning a way to use keyword-rich anchor text in your internal links. Avoid using text such as "More" or "Click here", and make sure the

technical and creative teams understand this. You also need to invest time in preparing an anchor text strategy for the site.

Use breadcrumb navigation

Breadcrumb navigation is a way to show the user where he is in the navigation hierarchy. Figure 4-4 shows an example from PetSmart.

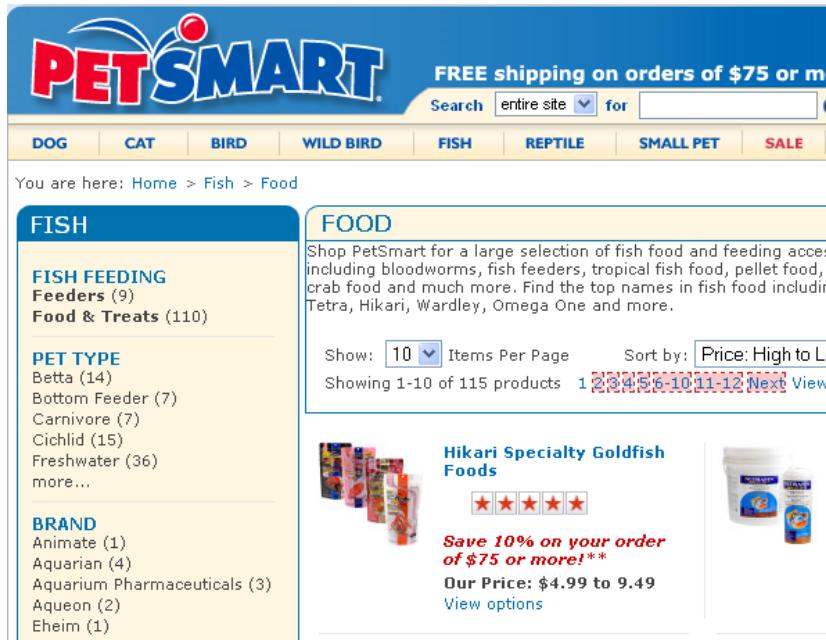


FIGURE 4-4. Breadcrumb bar on PetSmart.com

This page is currently two levels down from the home page. Also, note how the anchor text in the breadcrumb is keyword-rich, as is the menu navigation on the left. This is helpful to both users and search engines.

Minimize link depth

Search engines (and users) look to the site architecture for clues as to what pages are most important. A key factor is how many clicks from the home page it takes to reach a page. A page that is only one click from the home page is clearly important. A page that is five clicks away is not nearly as influential. In fact, the search engine spider may never even find such a page (depending in part on the site's link authority).

Standard SEO advice is to keep the site architecture as flat as possible, to minimize clicks from the home page to all important content. Do not go off the deep end, though; too many links

on a page are not good for search engines either (a standard recommendation on this is not to exceed 100 links from a web page; we will cover this in more detail in [Chapter 6](#)). The bottom line is that you need to plan out a site structure that is as flat as you can reasonably make it without compromising the user experience.

In this and the preceding sections, we outlined common structural decisions that you need to make prior to beginning your SEO project. There are other considerations, such as how you might be able to make your efforts scale across a very large site (thousands of pages or more). In such a situation, you cannot feasibly review every page one by one.

Auditing an Existing Site to Identify SEO Problems

Auditing an existing site is one of the most important tasks that SEO professionals encounter. SEO is a relatively new field, and many of the limitations of search engine crawlers are non-intuitive. In addition, web developers are generally not well versed in SEO. This includes those who have developed CMSs, so there is a lot of opportunity to find problems when conducting a site audit.

Elements of an Audit

As we will discuss in [Chapter 6](#), your website needs to be a strong foundation for the rest of your SEO efforts to succeed. An SEO site audit is often the first step in executing an SEO strategy.

The following sections identify what you should look for when performing a site audit.

Usability

Although this may not be seen as a direct SEO issue, it is a very good place to start. Usability affects many factors, including conversion rate, as well as the propensity of people to link to a site.

Accessibility/spiderability

Make sure the site is friendly to search engine spiders. We discuss this in detail in “[Making Your Site Accessible to Search Engines](#)” on page 181 and “[Creating an Optimal Information Architecture](#)” on page 187 in [Chapter 6](#).

Search engine health check

Here are some quick health checks:

- Perform a `site:yourdomain.com` search in the search engines to make sure all your pages appear to be in the index. Compare this to the number of unique pages you believe you have on your site.

- Test a search on your brand terms to make sure you are ranking for them (if not, you may be suffering from a penalty).
- Check the Google cache to make sure the cached versions of your pages look the same as the live versions of your pages.

Keyword health checks

Are the right keywords being targeted? Does the site architecture logically flow from the way users search on related keywords? Does more than one page target the same exact keyword (a.k.a. *keyword cannibalization*)? We will discuss these items in the section “[Keyword Targeting](#)” on page 211 in [Chapter 6](#).

Duplicate content checks

The first thing you should do is to make sure the non-www version of your pages (i.e., *http://yourdomain.com*) 301-redirects to the www version of your pages (i.e., *http://www.yourdomain.com*), or vice versa (this is often called the *canonical redirect*). While you are at it, check that you don’t have https: pages that are duplicates of your http: pages. You should check the rest of the content on the site as well.

The easiest way to do this is to take unique strings from each of the major content pages on the site and search on them in Google. Make sure you enclose the string inside double quotes (e.g., “a phrase from your website that you are using to check for duplicate content”) so that Google will search for that exact string.

If your site is monstrously large and this is too big a task, make sure you check the most important pages, and have a process for reviewing new content before it goes live on the site.

You can also use commands such as `inurl:` and `intitle:` to check for duplicate content. For example, if you have URLs for pages that have distinctive components to them (e.g., “1968-mustang-blue” or “1097495”), you can search for these with the `inurl:` command and see whether they return more than one page.

Another duplicate content task to perform is to make sure each piece of content is accessible at only one URL. This probably trips up more big, commercial sites than any other issue. The issue is that the same content is accessible in multiple ways and on multiple URLs, forcing the search engines (and visitors) to choose which is the canonical version, which to link to, and which to disregard. No one wins when sites fight themselves—make peace, and if you have to deliver the content in different ways, rely on cookies so that you don’t confuse the spiders.

URL check

Make sure you have clean, short, descriptive URLs. *Descriptive* means keyword-rich but not keyword-stuffed. You don’t want parameters appended (or have a minimal number if you must have any), and you want them to be simple and easy for users (and spiders) to understand.

Title tag review

Make sure the title tag on each page of the site is unique and descriptive. Ideally, don't waste your time (or limited space) by including the brand name of your organization in the URL. If you must include it, the brand name should show up at the end of the title tag, not at the beginning, as placement of keywords at the front of a URL brings ranking benefits. Also check to make sure the title tag is fewer than 70 characters long.

Content review

Do the main pages of the site have enough content? Do these pages all make use of header tags? A subtler variation of this is making sure the percentage of pages on the site with little content is not too high compared to the total number of pages on the site.

Meta tag review

Check for a meta robots tag on the pages of the site. If you find one, you may have already spotted trouble. An unintentional NoIndex or NoFollow tag (we define these in “[Content Delivery and Search Spider Control](#)” on page 238 in Chapter 6) could really mess up your search ranking plans.

Also make sure every page has a unique meta description. If for some reason that is not possible, consider removing the meta description altogether. Although the meta description tags are not a significant factor in ranking, they may well be used in duplicate content calculations, and the search engines frequently use them as the description for your web page in the SERPs; therefore, they affect click-through rate.

Sitemaps file and robots.txt file verification

Use the Google Webmaster Tools robots.txt verification tool to check your *robots.txt* file. Also verify that your Sitemaps file is identifying all of your (canonical) pages.

Redirect checks

Use a server header checker such as [Live HTTP Headers](#) to check that all the redirects used on the site return a 301 HTTP status code. Check all redirects this way to make sure the right thing is happening. This includes checking that the canonical redirect is properly implemented.

Unfortunately, given the non-intuitive nature of why the 301 is preferred, you should verify that this has been done properly even if you provided explicit direction to the web developer in advance. Mistakes do get made, and sometimes the CMS or the hosting company makes it difficult to use a 301.

Internal linking checks

Look for pages that have excess links. Google advises 100 per page as a maximum, although it is OK to go with more on more important and heavily linked-to pages.

Make sure the site makes good use of anchor text on its internal links. This is a free opportunity to inform users and search engines what the various pages of your site are about. Don't abuse it, though. For example, if you have a link to your home page in your global navigation (which you should), call it "Home" instead of picking your juiciest keyword. The anchor text of internal links to the home page is not helpful for rankings anyway.

The search engines view this particular practice as spammy, and it does not represent a good user experience. Keep using that usability filter through all of these checks!

N O T E

A brief aside about hoarding PageRank: many people have taken this to an extreme and built sites where they refused to link out to other quality websites, because they feared losing visitors and link juice. Ignore this idea! You should link out only to quality websites. It is good for users, and it is likely to bring ranking benefits (through building trust and relevance based on what sites you link to). Just think of the human user and deliver what he'd want. It is remarkable how far this will take you.

Avoidance of unnecessary subdomains

The engines may not apply the entirety of a domain's trust and link juice weight to subdomains. This is largely due to the fact that a subdomain could be under the control of a different party, and therefore in the search engine's eyes it needs to be separately evaluated. In the great majority of cases, subdomain content can easily go in a subfolder.

Geolocation

If the domain is targeting a specific country, make sure the guidelines for country geotargeting outlined in "[Best Practices for Multilanguage/Country Targeting](#)" on page 275 in [Chapter 6](#) are being followed. If your concern is more about ranking for *chicago pizza* because you own a pizza parlor in Chicago, make sure your address is on every page of your site. You should also check your results in Google Local to see whether you have a problem there.

External linking

Check the inbound links to the site. Use a backlinking tool such as [Yahoo! Site Explorer](#), [Linkscape](#), [Majestic-SEO](#), or [Link Diagnosis](#) to collect data about your links. Look for bad patterns in the anchor text, such as 87% of the links having the critical keyword for the site in them. Unless the critical keyword happens to also be the name of the company, this is a sure sign of trouble. This type of distribution is quite likely the result of purchasing links or other manipulative behavior.

On the flip side, make sure the site's critical keyword is showing up a fair number of times. A lack of the keyword usage in inbound anchor text is not good either. You need to find a balance.

Look to see that there are links to pages other than the home page. These are often called *deep links* and they will help drive the ranking of key sections of your site. You should also look at the links themselves. Visit the linking pages and see whether the links appear to be paid for. They may be overtly labeled as sponsored, or their placement may be such that they are clearly not a natural endorsement. Too many of these are another sure trouble sign.

Lastly on the topic of external links, make sure that there are enough links, and that there are also enough high-quality links in the mix. How does the link profile for the site compare to the link profiles of its major competitors?

Page load time

Is the page load time excessive? Too long a load time may slow down crawling and indexing of the site. However, to be a factor, this really needs to be excessive—certainly longer than five seconds, and perhaps even longer than that.

Image alt tags

Do all the images have relevant keyword-rich image alt attributes text and filenames? Search engines can't easily tell what is inside an image, and the best way to provide them with some clues is with the alt attribute and the filename of the image. These can also reinforce the overall context of the page itself.

Code quality

Although W3C validation is not something the search engines require, checking the code itself is a good idea. Poor coding can have some undesirable impacts. As we previously discussed, use a tool such as [SEO Browser](#) to see how the search engines see the page.

The Importance of Keyword Reviews

Another critical component of an architecture audit is a keyword review. Basically, this involves the following steps.

Step 1: Keyword research

It is vital to get this done as early as possible in any process. Keywords drive on-page SEO, so you want to know which ones to target. You can read about this in more detail in [Chapter 5](#).

Step 2: Site architecture

Coming up with a site architecture can be very tricky. At this stage, you need to look at your keyword research and the existing site (to make as few changes as possible). You can think of this in terms of your site map.

You need a hierarchy that leads you to each of your *money pages* (i.e., the pages where conversions are most likely to occur). Obviously, a good site hierarchy allows the parents of your money pages to rank for relevant keywords (which are likely to be shorter tail).

Most products have an obvious hierarchy they fit into, but when you start talking in terms of anything that naturally has multiple hierarchies, it gets incredibly tricky. The trickiest hierarchies, in our opinion, occur when there is a location involved. In London alone there are London boroughs, metropolitan boroughs, Tube stations, and postcodes. London even has a city (“The City of London”) within it.

In an ideal world, you will end up with a single hierarchy that is natural to your users and gives the closest mapping to your keywords. Whenever there are multiple ways in which people search for the same product, establishing a hierarchy becomes challenging.

Step 3: Keyword mapping

Once you have a list of keywords and a good sense of the overall architecture, start mapping the major relevant keywords to URLs (not the other way around). When you do this, it is a very easy job to spot pages that you were considering that aren’t targeting a keyword, and more importantly, keywords that don’t have a page.

It is worth pointing out that between step 2 and step 3 you will remove any wasted pages.

If this stage is causing you issues, revisit step 2. Your site architecture should lead naturally to a mapping that is easy to use and includes your keyphrases.

Step 4: Site review

Once you are armed with your keyword mapping, the rest of the site review becomes a lot easier. Now when you are looking at title tags and headings, you can refer back to your keyword mapping and not only see whether the heading is in an h1 tag, but also see whether it includes the right keywords.

Keyword Cannibalization

Keyword cannibalization typically starts when a website’s information architecture calls for the targeting of a single term or phrase on multiple pages of the site. Many times this is done unintentionally, but results in several or even dozens of pages that have the same keyword target in the title and header tags. [Figure 4-5](#) shows the problem.

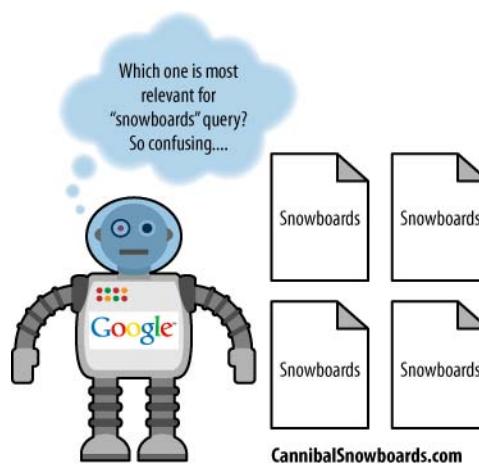


FIGURE 4-5. Example of keyword cannibalization

Search engines will spider the pages on your site and see 4 (or 40) different pages, all seemingly relevant to one particular keyword (in the example in [Figure 4-5](#), the keyword is *snowboards*). For clarity's sake, Google doesn't interpret this as meaning that your site as a whole is more relevant to snowboards or should rank higher than the competition. Instead, it forces Google to choose among the many versions and pick one it feels best fits the query. When this happens you lose out on a number of rank-boosting features:

Internal anchor text

Since you're pointing to so many different pages with the same subject, you can't concentrate the value of internal anchor text on one target.

External links

If four sites link to one page on snowboards, three sites link to another of your snowboard pages, and six sites link to yet another snowboard page, you've split up your external link value among three pages, rather than consolidating it into one.

Content quality

After three or four pages of writing about the same primary topic, the value of your content is going to suffer. You want the best possible single page to attract links and referrals, not a dozen bland, replicated pages.

Conversion rate

If one page is converting better than the others, it is a waste to have multiple, lower-converting versions targeting the same traffic. If you want to do conversion tracking, use a multiple-delivery testing system (either A/B or multivariate).

So, what's the solution? Take a look at [Figure 4-6](#).

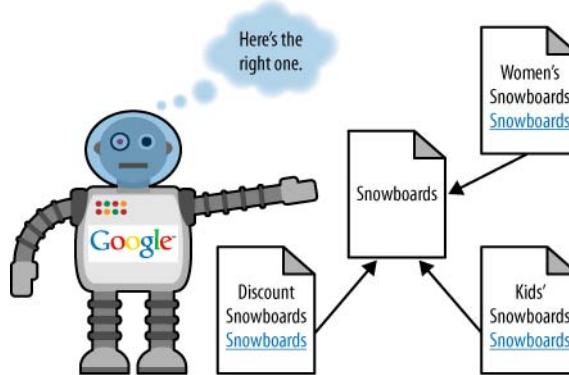


FIGURE 4-6. Solution to keyword cannibalization

The difference in this example is that instead of targeting the singular *snowboards* on every page, the pages are focused on unique, valuable variations and all of them link back to an original, canonical source for the singular term. Google can now easily identify the most relevant page for each of these queries. This isn't just valuable to the search engines; it is also a far better user experience and overall information architecture.

What should you do if you've already got a case of keyword cannibalization? Employ 301s liberally to eliminate pages competing with each other, or figure out how to differentiate them. Start by identifying all the pages in the architecture with this issue and determine the best page to point them to, and then use a 301 from each of the problem pages to the page you wish to retain. This ensures not only that visitors arrive at the right page, but also that the link equity and relevance built up over time are directing the engines to the most relevant and highest-ranking-potential page for the query.

Example: Fixing an Internal Linking Problem

Enterprise sites range between 10,000 and 10 million pages in size. For many of these types of sites, an inaccurate distribution of internal link juice is a significant problem. Figure 4-7 shows how this can happen.

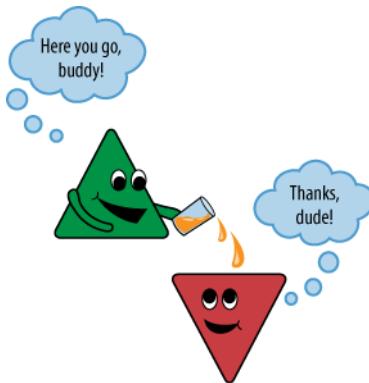


FIGURE 4-8. Using cross-links to push link juice where you want it

Enterprise-size domain

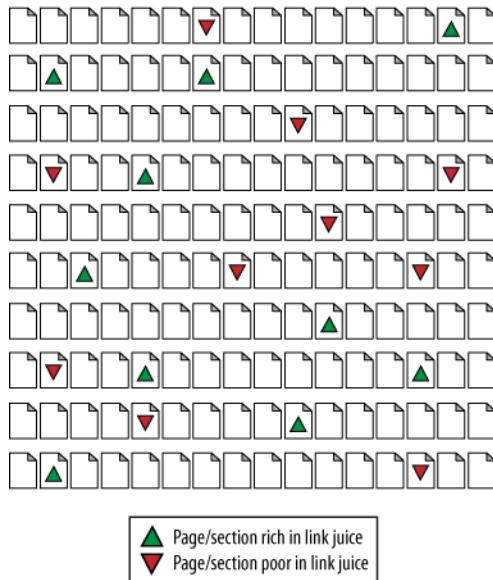


FIGURE 4-7. Link juice distribution on a very large site

Figure 4-7 is an illustration of the link juice distribution issue. Imagine that each of the tiny pages represents between 5,000 and 100,000 pages in an enterprise site. Some areas, such as blogs, articles, tools, popular news stories, and so on, might be receiving more than their fair share of internal link attention. Other areas—often business-centric and sales-centric content—tend to fall by the wayside. How do you fix it? Take a look at Figure 4-8.

The solution is simple, at least in principle. Have the link-rich pages spread the wealth to their link-bereft brethren. As easy as this looks, in execution it can be incredibly complex. Inside the architecture of a site with several hundred thousand or a million pages, it can be nearly impossible to identify link-rich and link-poor pages, never mind adding code that helps to distribute link juice equitably.

The answer, sadly, is labor-intensive from a programming standpoint. Enterprise site owners need to develop systems to track inbound links and/or rankings and build bridges (or, to be more consistent with [Figure 4-8](#), spouts) that funnel juice between the link-rich and link-poor.

An alternative is simply to build a very flat site architecture that relies on relevance or semantic analysis (several enterprise-focused site search and architecture firms offer these). This strategy is more in line with the search engines' guidelines (though slightly less perfect) and is certainly far less labor-intensive.

Interestingly, the rise of massive weight given to domain authority over the past two to three years appears to be an attempt by the search engines to overrule potentially poor internal link structures (as designing websites for PageRank flow really doesn't serve users particularly well), and to reward sites that have massive authority, inbound links, and trust.

Server and Hosting Issues

Thankfully, few server or web hosting dilemmas affect the practice of search engine optimization. However, when overlooked, they can spiral into massive problems, and so are worthy of review. The following are server and hosting issues that can negatively impact search engine rankings:

Server timeouts

If a search engine makes a page request that isn't served within the bot's time limit (or that produces a server timeout response), your pages may not make it into the index at all, and will almost certainly rank very poorly (as no indexable text content has been found).

Slow response times

Although this is not as damaging as server timeouts, it still presents a potential issue. Not only will crawlers be less likely to wait for your pages to load, but surfers and potential linkers may choose to visit and link to other resources because accessing your site becomes a problem.

Shared IP addresses

Lisa Barone wrote [an excellent post](#) on the topic of shared IP addresses back in March 2007. Basic concerns include speed, the potential for having spammy or untrusted neighbors sharing your IP address, and potential concerns about receiving the full benefit of links to your IP address (discussed in more detail at <http://www.seroundtable.com/archives/002358.html>).

Blocked IP addresses

As search engines crawl the Web, they frequently find entire blocks of IP addresses filled with nothing but egregious web spam. Rather than blocking each individual site, engines do occasionally take the added measure of blocking an IP address or even an IP range. If you're concerned, search for your IP address at Bing using the *IP:address* query.

Bot detection and handling

Some sys admins will go a bit overboard with protection and restrict access to files to any single visitor making more than a certain number of requests in a given time frame. This can be disastrous for search engine traffic, as it will constantly limit the spiders' crawling ability.

Bandwidth and transfer limitations

Many servers have set limitations on the amount of traffic that can run through to the site. This can be potentially disastrous when content on your site becomes very popular and your host shuts off access. Not only are potential linkers prevented from seeing (and thus linking to) your work, but search engines are also cut off from spidering.

Server geography

This isn't necessarily a problem, but it is good to be aware that search engines do use the location of the web server when determining where a site's content is relevant from a local search perspective. Since local search is a major part of many sites' campaigns and it is estimated that close to 40% of all queries have some local search intent, it is very wise to host in the country (it is not necessary to get more granular) where your content is most relevant.

Identifying Current Server Statistics Software and Gaining Access

In [Chapter 9](#), we will discuss in detail the methods for tracking results and measuring success, and we will also delve into how to set a baseline of measurements for your SEO projects. But before we do, and before you can accomplish these tasks, you need to have the right measurement systems in place.

Web Analytics

Analytics software can provide you with a rich array of valuable data about what is taking place on the site. It can answer questions such as:

- How many unique visitors did you receive yesterday?
- Is traffic trending up or down?
- What are the most popular search terms with which people find you?
- What are the most popular pages on your site?

- What are the best converting pages on the site?

We strongly recommend that if your site does not have any measurement systems in place currently, you put something in place immediately. High-quality, free analytics tools are available, such as [Yahoo! Web Analytics](#) and [Google Analytics](#). Of course, higher-end analytics solutions are also available, and we will discuss them in more detail in [Chapter 9](#).

LogFile Tracking

Logfiles contain a detailed click-by-click history of all requests to your web server. Make sure you have access to the logfiles and some method for analyzing them. If you use a third-party hosting company for your site, chances are it provides some sort of free logfile analyzer, such as [AWStats](#), [Webalizer](#), or something similar. Obtain access to whatever tool is in use as soon as you can.

What these tools do that JavaScript-based web analytics software cannot is record search engine spider activity on your site. Although spidering will typically vary greatly from day to day, you can still see longer-term trends of search engine crawling patterns, and whether crawling activity is trending up (good) or down (bad). So, although this web crawler data is very valuable, do not rely on these free solutions provided by hosting companies for all of your analytics data, as there is a lot of value in what traditional analytics tools can capture.

N O T E

Some web analytics software packages read logfiles as well, and therefore can report on crawling activity. We will discuss these in more detail in [Chapter 9](#).

Google and Bing Webmaster Tools

Two other valuable sources of data are [Google Webmaster Tools](#) and [Bing Webmaster Tools](#). We cover these extensively in “[Using Search-Engine-Supplied SEO Tools](#)” on page 487 in [Chapter 11](#).

From a planning perspective, you will want to get these tools in place as soon as possible. Both tools provide valuable insight into how the search engines see your site. This includes things such as external link data, internal link data, crawl errors, high-volume search terms, and much, much more.

N O T E

Some companies will not want to set up these tools because they do not want to share their data with the search engines, but this is a nonissue as the tools do not provide the search engines with any more data about your website; rather, they let you see some of the data the search engines already have.

Search Analytics

Search analytics is a new and emerging category of tools. Search analytics tools specifically monitor how your website interacts with the search engines. The company most known for this (as of early 2009) is [Enquisite](#). In [Chapter 5](#), we will cover how you can use Enquisite Optimizer to help you find long tail keywords that you can use to optimize your site content or your PPC campaigns.

Although this category is in its infancy, it is worth monitoring closely to see what tools become available that can provide your organization with an advantage in competing for search traffic.

Determining Top Competitors

Understanding the competition should be a key component of planning your SEO. The first step is to understand who your competitors in the search results really are. It can often be small players who give you a run for your money. For example, consider the previously mentioned credit card search in Google; Visa, Master Card, American Express, and Discover Card all fail to reach the #1 position in the Google results.

Instead, you will find these results dominated by affiliate players. Affiliates tend to be the most adept at search engine optimization and can be the most lax in abiding by the search engines' terms and conditions.

Two Spam Examples

Affiliates that cheat tend to come and go out of the top search results, as only sites that implement ethical tactics are likely to maintain their positions over time. You can help expedite the cheaters' fall from grace by reporting them to Google at <http://www.google.com/contact/spamreport.html>, or better yet, via the dashboard in your Google Webmaster Tools account (where it will carry more weight).

How do you know whether a top-ranking site is playing by the rules? Look for dubious links to the site using a backlink analysis tool such as [Linkscape](#). Since the number of links is one factor search engines use to determine search position, less ethical websites will obtain links from a multitude of irrelevant and low-quality sites.

This sort of sleuthing can reveal some surprises. For instance, here are examples of two devious link schemes:

- GiftCertificates.com's short-lived nemesis was [FindGiftCards.com](#), which came out of nowhere to command the top two spots in Google for the all-important search term *gift certificates*, thus relegating [GiftCertificates.com](#) to the third position. How did [FindGiftCards.com](#) do it? It operated a sister site, [123counters.com](#), with a free hit counter that propagated "link spam" across thousands of sites, all linking back to [FindGiftCards.com](#) and other sites in its network.

Sadly for [FindGiftCards.com](#), Stephan Spencer, founder and president of the e-marketing agency Netconcepts, outed the company in an article he wrote for *Multichannel Merchant* back in 2004 (http://multichannelmerchant.com/catalogue/ar/marketing_casing_competition/), and Google became aware of the scam. The end result? Down to two pages in the Google index, as shown in [Figure 4-9](#).

- [CraigPadoa.com](#) was a thorn in the side of [SharperImage.com](#), outranking the latter for its most popular product, the Ionic Breeze, by frameset trickery and guestbook spamming (in other words, defacing vulnerable websites with fake guestbook entries that contained spammy links back to its own site). As soon as The Sharper Image realized what was happening, it jumped on the wayward affiliate. It also restricted such practices in its affiliate agreement and stepped up its monitoring for these spam practices.

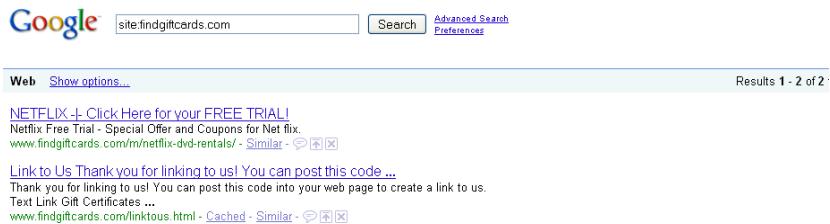


FIGURE 4-9. Site with only two pages in the index

Seeking the Best

Look for competitors whose efforts you would like to emulate (or “embrace and extend,” as Bill Gates would put it)—usually a website that consistently dominates the upper half of the first page of search results in the search engines for a range of important keywords that are popular and relevant to your target audience. Note that your “mentor” competitors shouldn’t just be good performers; they should also demonstrate that they know what they’re doing when it comes to SEO. To assess competitors’ competence at SEO, you need to answer the following questions:

- Are their websites fully indexed by Google and Yahoo!? In other words, are all their web pages, including product pages, making it into the search engines’ databases? You can use the free tool at <http://www.netconcepts.com/urlcheck> to find out, or you can go to each search engine and type in `site:theirdomain.com`. A competitor with only a small percentage of its site in Google probably has a site that is unfriendly to search spiders.
- Do their product and category pages have keyword-rich page titles (title tags) unique to each page? You can easily review an entire site’s page titles within Google or Yahoo! by searching for `site:www.yourcompetitor.com`.

Incidentally, this type of search can sometimes yield confidential information. A lot of webmasters do not realize that Google has discovered and indexed commercially sensitive

content buried deep in its site. For example, a Google search for *confidential business plan filetype:doc* will yield a lot of real business plans among the sample templates.

- Do their product and category pages have reasonably high PageRank scores?
- Is anchor text across the site, particularly in the navigation, keyword-rich?
- Are the websites getting penalized? You can overdo SEO. Too much keyword repetition or too many suspiciously well-optimized text links can yield a penalty for over-optimization. Sites can also be penalized for extensive amounts of duplicate content. You can learn more about how to identify search engine penalties in the section “[Content Theft](#)” on page 467 in [Chapter 10](#).
- Are they spamming the search engines with “doorway pages”?

Uncovering Their Secrets

Let’s assume your investigation has left you with several competitors who are gaining excellent search placement using legitimate, intelligent tactics. Now it is time to uncover their secrets:

- What keywords are they targeting? You can determine this by looking at the page titles (up in the blue bar at the top of your web browser, which also appears in the search results listings) across their home page and product category pages, then by looking at their meta keywords tag (right-click, select View Source, and then scour the HTML source for the list of keywords that follow the bit of HTML that looks something like the following:

```
<meta name="keywords" content="keyword1, keyword2, ..." >
```

- Who’s linking to their home page, or to their top-selling product pages and category pages? A link popularity checker can be quite helpful in analyzing this. Amazon.com has been especially successful in achieving deep links into its millions of products.
- If it is a database-driven site, what technology tricks are they using to get search engine spiders such as Googlebot to cope with the site being dynamic? Nearly all the technology tricks are tied to the e-commerce platforms the competitors are running. You can check to see whether they are on the same server software as you by using the “What’s that site running?” tool on the top-left corner of <http://www.netcraft.com>. [Figure 4-10](#) shows a screen shot of the results for [HSN.com](#).

While you are at it, look at “cached” (archived) versions of your competitors’ pages by clicking on the Cached link next to their search results in Google to see whether they’re doing anything too aggressive, such as *cloaking*, where they serve up a different version of the page to search engine spiders than to human visitors. The cached page will show you what the search engine actually saw, and you can see how it differs from the page you see when you go to the web page yourself.

- What effect will their future SEO initiatives have on their site traffic? Assess the success of their SEO not just by the lift in rankings. Periodically record key SEO metrics over time—

OS, Web Server and Hosting History for hsn.com					
http://hsn.com was running BigIP on F5 Big-IP when last queried at 22-Jul-2009 14:13:44 GMT - refresh now Site Report FAQ Try out the Netcraft Toolbar!					
OS	Server	Last changed	IP address	Netblock	Owner
F5 Big-IP	BigIP	22-Jul-2009	192.234.237.190		Home Shopping Network, Inc.
F5 Big-IP	BigIP	18-Apr-2009	192.234.237.190		Home Shopping Network, Inc.
F5 Big-IP	BigIP	17-Apr-2009	161.254.200.190		Home Shopping Network
F5 Big-IP	BigIP	16-Apr-2009	192.234.237.190		Home Shopping Network, Inc.
F5 Big-IP	Big-IP	27-Feb-2009	192.234.237.190		Home Shopping Network, Inc.
F5 Big-IP	Big-IP	26-Feb-2009	161.254.200.190		Home Shopping Network
F5 Big-IP	Big-IP	25-Feb-2009	192.234.237.190		Home Shopping Network
F5 Big-IP	Big-IP	5-Nov-2008	161.254.200.190		Home Shopping Network
F5 Big-IP	Big-IP	3-Nov-2008	192.234.237.190		Home Shopping Network, Inc.
F5 Big-IP	Big-IP	1-Nov-2008	161.254.200.190		Home Shopping Network

FIGURE 4-10. Sample Netcraft output

the number of pages indexed, the PageRank score, the number of links—and watch the resulting effect on their site traffic.

You do not need access to competitors' analytics data or server logs to get an idea of how much traffic they are getting. Simply go to [Compete](#), [Quantcast](#), or [Alexa](#), and search on the competitor's domain. If you have the budget for higher-end competitive intelligence tools, you can use [Compete](#) or [Hitwise](#).

The data these tools can provide is limited in its accuracy, but still very useful in giving you a general assessment of where they are. The tools are most useful when making relative comparisons between sites in the same market space. To get an even better idea, use their capabilities to compare the traffic of multiple sites. In this mode, you can get a pretty accurate idea as to how your traffic compares to theirs.

You can now get this type of data directly from Google as well, using [Google Trends for Websites](#). The output of this tool is just a summary of Google traffic, but it is a much larger data set than is available from the other products. Figure 4-11 shows an example of the output from Google Trends for Websites.

Note that tools such as Alexa, Compete, and Quantcast do have other unique features and functionality not available in Google Trends for Websites.

- How does the current state of their sites' SEO compare with those of years past? You can reach back into history and access previous versions of your competitors' home pages and view the HTML source to see which optimization tactics they were employing back then. The [Wayback Machine](#) provides an amazingly extensive archive of web pages.

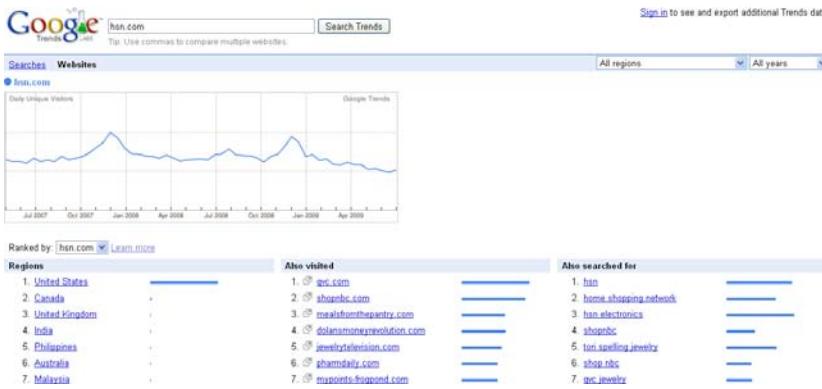


FIGURE 4-11. Google Trends for Websites

Assessing Historical Progress

Measuring the results of SEO changes can be challenging, partly because there are so many moving parts and partly because months can elapse between when changes are made to a site and when results are seen in search rankings and traffic. This difficulty only increases the importance of measuring progress and being accountable for results. This section will explore methods for measuring the results from your SEO efforts.

Maintain a Timeline of Site Changes

Keeping a log of changes to your site is highly recommended. If you're not keeping a timeline (which could be as simple as an online spreadsheet or as complex as a professional project management visual flowchart), you will have a harder time executing your SEO plan. Sure, without one you can still gauge the immediate effects of content additions/revisions, link acquisitions, and development changes, but there's obscured visibility into what technical modifications to the website might have altered the course of search traffic, either positively or negatively.

If you can't map changes, both those intended to influence SEO and those for which SEO wasn't even a consideration, you'll be optimizing blind and could miss powerful signals that could help dictate your strategy going forward. There are also many scenarios in which you will want to try to establish cause and effect, such as:

If search traffic spikes or plummets

Sudden changes in organic traffic are obviously notable events. If traffic plummets, you will be facing lots of questions about why, and having a log of site changes will put you in a better position to address whether some changes you recommended could have been

the cause. Of course, if traffic spikes you will want to be able to see whether an SEO-related change was responsible as well.

When gradual traffic changes begin

Changes do not always come as sudden spikes or drop-offs. Even so, when you see the traffic beginning a gradual climb you will want to be able to assess the likely reasons.

To track and report SEO progress

Accountability is a key component of SEO. Budget managers will want to know what return they are getting on their SEO investment. This will inevitably fall into two buckets: itemizing specific work items worked on; and benefits to the business. An ongoing change log makes all of this much easier to accomplish.

Types of Site Changes That Can Affect SEO

Your log should track all changes to the website, not just those that were made with SEO in mind. Organizations make many changes that they do not think will affect SEO, but they have a big impact on it. Here are some examples:

- Content areas/features/options added to the site (this could be anything from a new blog to a new categorization system).
- Changing the domain of your site. This can have a significant impact, and you should document when the switchover was made.
- Modifications to URL structures. Changes to URLs on your site will likely impact your rankings, so record any and all changes.
- Implementing a new CMS. This is a big one, with a very big impact. If you must change your CMS, make sure you do a thorough analysis of the SEO shortcomings of the new CMS versus the old one, and make sure you track the timing and the impact.
- New partnerships that either send links or require them (meaning your site is earning new links or linking out to new places).
- Changes to navigation/menu systems (moving links around on pages, creating new link systems, etc.).
- Any redirects, either to or from the site.
- Upticks in usage/traffic and the source (e.g., if you get mentioned in the press and receive an influx of traffic from it).

When you track these items, you can create an accurate storyline to help correlate causes with effects. If, for example, you've observed a spike in traffic from Yahoo! that started four to five days after you switched from menu links in the footer to the header, it is a likely indicator of a causal relationship.

Without such documentation it could be months before you notice the surge—and there would be no way to trace it back to the responsible modification. Your design team might later choose

to switch back to footer links, your traffic may fall, and no record would exist to help you understand why. Without the lessons of history, you are doomed to repeat the same mistakes.

Previous SEO Work

When you are brought on to handle the SEO for a particular website, one of the first things you need to find out is which SEO activities have previously been attempted. First of all, there may be valuable data there, such as a log of changes which you can match up with analytics data to gauge impact.

If no such log exists, you can always check the Wayback Machine (<http://www.archive.org>) to see whether it has historical logs for your website. This offers snapshots of what the site looked like at various points in time.

Even if a log was not kept, spend some time building a timeline of when any of the types of changes that affect SEO (as we discussed earlier in this section) took place. In particular, see whether you can get copies of the exact recommendations the prior SEO consultant made, as this will help you with the timeline and the specifics of the changes made.

You should also pay particular attention to understanding the types of link-building activities that took place. Were shady practices used that carry a lot of risk? Was there a particular link-building tactic that worked quite well? Going through the history of the link-building efforts can yield tons of information that you can use to determine next steps.

Benchmarking Current Indexing Status

The search engines have an enormous task—that of indexing the world’s online content; well, more or less. The reality is that they try hard to discover all of it, but they do not choose to include all of it in their index. This can happen for a variety of reasons, such as the page being inaccessible to the spider, being penalized, or not having enough link juice to merit inclusion.

When you launch a new site, or add new sections to an existing site, or are dealing with a very large site, not every page will necessarily make it into the index. To get a handle on this you will want to actively track the indexing level of your site. If your site is not fully indexed, it could be a sign of a problem (not enough links, poor site structure, etc.).

Getting basic indexation data from search engines is pretty easy. All three major search engines support the same basic syntax for that, which is `site:yourdomain.com`. [Figure 4-12](#) shows a sample of the output from Bing.

The screenshot shows the Bing search interface. In the top left is the Bing logo with a city skyline background. To its right is a search bar containing the query "site:alchemistmedia.com". Next to the search bar is a magnifying glass icon. On the far right of the header, it says "1-10 of 112 results · Advanced".

ALL RESULTS

SEARCH HISTORY
site:alchemistmedia.com
[See all](#)
[Clear all](#) | [Turn off](#)

San Francisco SEO Company Alchemist Media
San Francisco SEO Company Alchemist Media is a full service Search Engine Marketing consultancy with offices in San Francisco and Los Angeles, California, offering organic search ...
www.alchemistmedia.com · [Cached page](#)

Alchemy Of Search Engine Marketing
Hi folks, There hasn't been much blog activity in recent months because most of the focus has been on writing The Art of SEO book with Rand, Stephan, and Eric, for O'Reilly ...
www.alchemistmedia.com/blog · [Cached page](#)

Los Angeles SEO Company Alchemist Media
Los Angeles based Search Engine Optimization Company Alchemist Media is a full service SEO Consultancy with offices in Los Angeles and the Bay Area.
www.alchemistmedia.com/alchemist.htm · [Cached page](#)

San Francisco & Los Angeles SEO Company - Alchemist Media, Inc.
Search Engine Optimization Company Alchemist Media is a full service Search Engine Marketing Company, offering SEO and CPC services.
www.alchemistmedia.com/CPC_Click_Fraud.htm · [Cached page](#)

Alchemist Media SEO Audit Services & SEO Consulting
SEO Consulting Services including SEO audits, full SEO management and hourly SEO consulting on search engine optimization - Los Angeles & San Francisco.
www.alchemistmedia.com/seo/services.htm · [Cached page](#)

FIGURE 4-12. Indexing data from Bing

Keeping a log of the level of indexation over time can help you understand how things are progressing. This can take the form of a simple spreadsheet.

Related to indexation is the crawl rate of the site. Google provides this data in Google Webmaster Central. [Figure 4-13](#) shows a screen shot representative of the crawl rate charts that are available (another chart, not shown here, displays the average time spent downloading a page on your site).

Short-term spikes are not a cause for concern, nor are periodic drops in levels of crawling. What is important is the general trend. In [Figure 4-13](#), the crawl rate seems to be drifting upward. This bodes well for both rankings and indexation.

For the other search engines, the crawl-related data can then be revealed using logfile analyzers (see “[Auditing an Existing Site to Identify SEO Problems](#)” on page 106), and then a similar time line can be created and monitored.

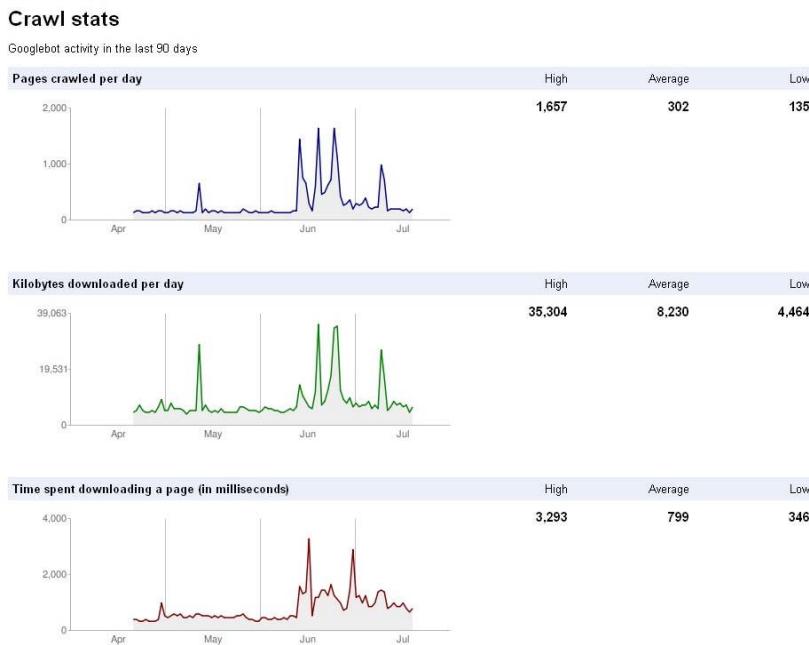


FIGURE 4-13. Crawl data from Google Webmaster Tools

Benchmarking Current Rankings

People really love to check their search rankings. Many companies want to use this as a measurement of SEO progress over time, but it is a bit problematic for a variety of reasons. Here is a summary of the major problems with rank checking:

- Google results are not consistent:
 - Different geographies (even in different cities within the United States) often give different results.
 - Different data centers give different results (and you can hit multiple data centers from a single location at different times).
 - Results are personalized for logged-in users based on their search histories.
 - No rank checker can monitor and report all of these inconsistencies (at least, not without scraping Google hundreds of times from all over the world with every possible setting).
- The Google API rarely matches up to what anyone sees in the search results:
 - It appears to match up only on very heavily trafficked, consistent search results—anything mid-tail or long tail is invariably inaccurate.

- It is extremely slow to update, so even though news results or geographic results might be mixed in (or even new sites or pages that have a large amount of recent link growth), the API won't update for days or sometimes weeks.
- Obsessing over rankings (rather than traffic) can result in poor strategic decisions:
 - When sites obsess over rankings for particular keywords, they often spend much more time and energy on a few particular keyphrases that produce far less value than if they had spent those resources on the site overall.
 - Long tail traffic is very often 70% to 80% of the demand, and it is much easier to rank in the long tail and get valuable traffic from the long tail than to concentrate on the few rankings at the top of the demand curve.

So, indulge your desire to check rankings by going to the search engine and typing in a few queries, but be sure to also keep an eye on your visitor and conversion statistics.

Benchmarking Current Traffic Sources and Volume

The most fundamental objective of any SEO project should be to drive the bottom line. For a business this means delivering more revenue with favorable ROI. As a predecessor to determining the level of ROI impact, the SEO practitioner must focus on increasing the volume of relevant traffic to the site. This is a more important objective than anything related to rankings or number of links obtained. More relevant traffic should mean more revenue for the business (or more conversions for those whose websites are not specifically selling something).

Today's web analytics tools make the gathering of such data incredibly easy. Two high-quality solutions are available that are completely free: Google Analytics (<http://www.google.com/analytics>) and Yahoo! Web Analytics (<http://web.analytics.yahoo.com/>). Figure 4-14 shows a sample of basic "unique visitors" data from Yahoo! Web Analytics.

Digging a little deeper, you can see the sources of the traffic as well, as you can see in Figure 4-15, which shows a Google Analytics report.

As an SEO practitioner, it will be natural to want to delve into more detail—specifically, to break down the search engine traffic and understand that better as well. Once again, this is easy to do in both tools (and in most of the commercially available tools out there), as shown in the Google Analytics screen shot in Figure 4-16.



FIGURE 4-14. “Unique visitors” report from Yahoo! Web Analytics

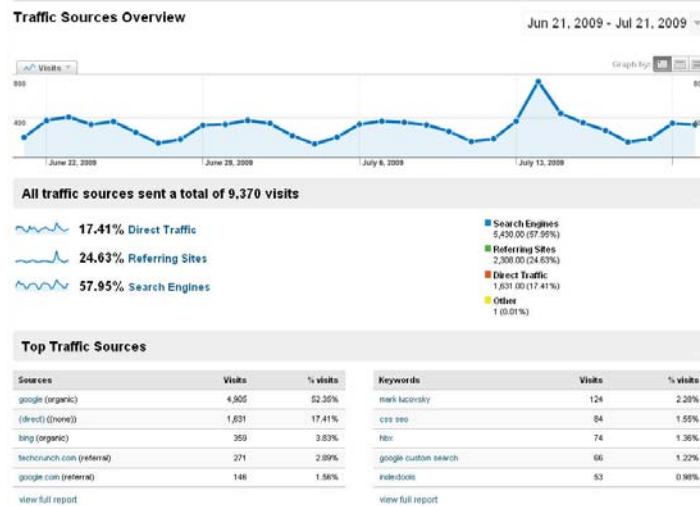


FIGURE 4-15. “Traffic sources” report from Google Analytics



FIGURE 4-16. “Search traffic” report from Google Analytics

This type of data allows you to see which search engines are delivering the majority of the traffic, and perhaps flag potential problems. In Figure 4-16, Yahoo! and Bing seem unusually low for this site.

Also, over on the right of [Figure 4-16](#) you can see that this site has an unusually high bounce rate. The site owner should probably work on some things to make the site a bit stickier.

Yet another thing to look at is which pages are getting the most traffic. [Figure 4-17](#) shows a sample report on that from Yahoo! Web Analytics.

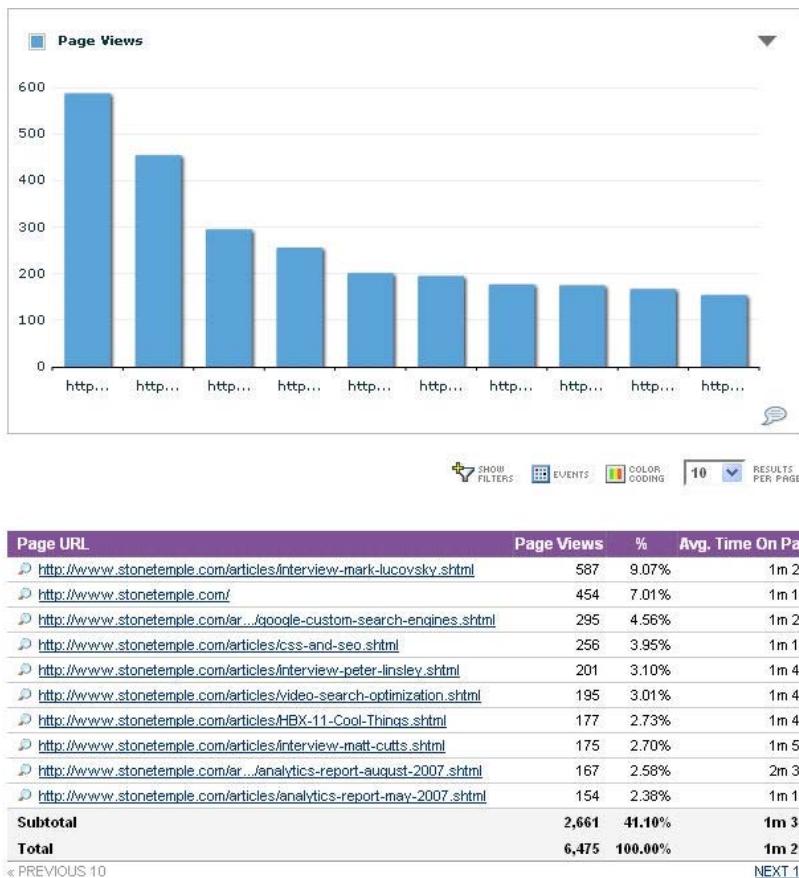


FIGURE 4-17. Yahoo! Web Analytics “most requested pages” report

In fact, the number of things you can look at in analytics is nearly endless. It is fair to say that there is too much data, and one of the key things that an SEO expert needs to learn is what data is worth looking at and what data is not worth looking at.

Leveraging Business Assets for SEO

Chances are your company/organization has a lot of valuable commodities beyond your website that can be put to good use to improve the quality and quantity of traffic you receive through search engine optimization efforts. We discuss some of these things in the subsections that follow.

Other Domains You Own/Control

If you have multiple domains, the major items to think about are:

- Can you 301-redirect some of those domains back to your main domain or to a subfolder on the site for additional benefit?
- Do you own exact keyword match domain names that would make for effective microsites?
- If you're maintaining those domains as separate sites, are you linking between them intelligently?

If any of those produce valuable strategies, pursue them—remember that it is often far easier to optimize what you're already doing than to develop entirely new strategies, content, and processes. Particularly on the link-building side, this is some of the lowest hanging fruit around.

Partnerships On and Off the Web

Partnerships can be leveraged in similar ways, particularly on the link-building front. If you have business partners that you supply or otherwise work with—or from whom you receive service—chances are good that you can implement link strategies between their sites and yours. Although reciprocal linking carries a bit of a bad reputation, there is nothing wrong with building a “partners,” “clients,” “suppliers,” or “recommended” list on your site, or with requesting that your organizational brethren do likewise for you. Just do this in moderation and make sure you link only to highly relevant, trusted sites.

Content or Data You've Never Put Online

Chances are that you have content that you never published on your website. This content can be immensely valuable to your SEO efforts. However, many companies are not savvy to the nuances of publishing that content in a manner that is friendly to search engines. Those hundreds of lengthy articles you published when you were shipping a print publication via the mail are a great fit for your website archives. You should take all of your email newsletters and make them accessible on your site. If you have unique data sets or written material, you should apply it to relevant pages on your site (or consider building out if nothing yet exists).

Customers Who Have Had a Positive Experience

Customers are a terrific resource for earning links, but did you also know they can write? Customers and website visitors can contribute all kinds of content. Seriously, if you have user-generated content (UGC) options available to you and you see value in the content your users produce, by all means reach out to customers, visitors, and email list subscribers for both links and content opportunities.

Your Fans

This principle applies equally to generic enthusiasts of your work. For many businesses that operate offline or work in entertainment, hard goods, or any consumer services, chances are good that if your business is worth its salt, you've got people who've used your products or services and would love to share their experience. Do you make video games? Reach out to your raving fans. Written a book? Mobilize your literary customers on the Web. Organize events? Like customers, fans are terrific resources for link acquisition, content creation, positive testimonials, and social media marketing (to help spread the word).

Combining Business Assets and Historical Data to Conduct SEO/Website SWOT Analysis

A classic staple of business school is the SWOT analysis—identifying strengths, weaknesses, opportunities, and threats faced by a business or project. By combining data from your business asset assessment and historical tracking data (and visitor analytics), you can create some very compelling analyses of your organization and its marketplace.

Identifying *strengths* is typically one of the easier objectives:

- What sources of traffic are working well for your site/business?
- Which projects/properties/partnerships are driving positive momentum toward traffic/revenue goals?
- Which of your content sections/types produces high traffic and ROI?
- What changes have you made historically that produced significant value?

Sourcing out the *weaknesses* can be tougher (and takes more intellectual honesty and courage):

- What content is currently sending low levels of search/visitor traffic?
- Which changes that were intended to produce positive results have shown little/no value?
- Which traffic sources are underperforming or underdelivering?
- What projects/properties/partnerships are being leveraged poorly?

Parsing *opportunities* requires a combination of strength and weakness analysis. You want to find areas that are doing well but have room to expand, as well as those that have yet to be explored:

- What brainstormed but undeveloped or untested projects/ideas can have a significant, positive impact?
- What traffic sources currently send good-quality traffic that could be expanded to provide more value?
- What areas of weakness have direct paths to recovery?
- Which website changes have had positive results? Can these be applied to other areas or applied more rigorously for increased benefit?
- What new markets or new content areas are potentially viable/valuable for expansion?
- What sources of new content/new links have yet to be tapped?

Determining *threats* can be the most challenging of the tasks. You'll need to combine creative thinking with an honest assessment of your weaknesses and your competitors' strengths, and consider the possibilities of macro-events that could shape your website/company's future:

- In your areas of weakness, which players in your market (or other, similar markets) are strong? How have they accomplished this?
- What shifts in human behavior, web usage, or market conditions could dramatically impact your business/site? (For example, consider the "what if people stopped searching and instead navigated the Web in different ways" perspective. It is a bit "pie in the sky," but it can make a big difference when, say, Expedia destroys the travel agency business or Craigslist makes classifieds obsolete.)
- Which competitors have had the most success in your arena? How have they accomplished this? Where does it intersect with your business/customers?
- Are there any strategies implemented by start-ups in similar businesses that have had massive success in a particular arena that could be dangerous to your business if they were replicated in your market?

Conducting SWOT analysis from a web marketing and SEO perspective is certainly one of the most valuable first steps you can take as an organization poised to expend resources. If you haven't taken the time to analyze from these bird's-eye-view perspectives, you might end up like a great runner who's simply gone off the course—sure, you'll finish fast, but where will it take you?

Conclusion

The first steps of SEO can often be challenging ones. It is also tempting to launch into the effort just to get things moving. However, spending some focused effort on setting up the SEO strategy before implementation will pay big dividends in the long run. Establish a strong foundation and you will help set yourself up for SEO success.

CHAPTER FIVE

Keyword Research

The Theory Behind Keyword Research

KEYWORD RESEARCH IS ONE OF THE MOST IMPORTANT, VALUABLE, AND HIGH-RETURN activities in the search engine marketing field. Through the detective work of dissecting your market's keyword demand, you learn not only which terms and phrases to target with SEO, but also more about your customers as a whole.

With keyword research you can predict shifts in demand, respond to changing market conditions, and produce the products, services, and content that web searchers are already actively seeking. In the history of marketing, there has never been such a low barrier to entry in understanding the motivations of consumers in virtually every niche.

Every search phrase that's typed into an engine is recorded in one way or another, and keyword research tools such as the ones we discuss in this chapter allow you to retrieve this information. However, those tools cannot show you (directly) how valuable or important it might be to rank for and receive traffic from those searches. To understand the value of a keyword, you need to research further, make some hypotheses, test, and iterate—the classic web marketing formula. This chapter seeks to expose the details of this process and the tools that can best assist.

Understanding the Long Tail of the Keyword Demand Curve

It is wonderful to deal with keywords that have 5,000 searches per day, or even 500 searches per day, but in reality these “popular” search terms actually comprise less than 30% of the

overall searches performed on the Web. The remaining 70% lie in what's commonly called the "long tail" of search (according to Enquisite and as published at <http://www.seomoz.org/blog/rewriting-the-beginners-guide-part-v-keyword-research>); see Figure 5-1. The tail contains hundreds of millions of unique searches that might be conducted a few times in any given day, or even only once ever, but when assessed in aggregate they comprise the majority of the world's demand for information through search engines.

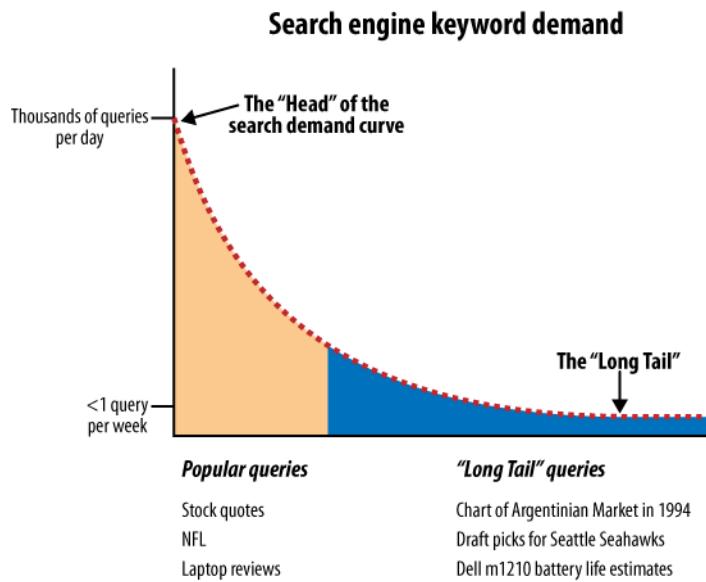


FIGURE 5-1. Long tail of search

Understanding the search demand curve is critical, because it stresses the importance of long-tail-targeted content—pages with information that is not directed at any particular single, popular query, but rather that captures a much broader range of the types of queries users enter into search engines.

Traditional Approaches: Domain Expertise, Site Content Analysis

One of the smartest things you can do when initially conducting keyword research is to use automated tools to brainstorm original ideas with the participants in the business. This can be surprisingly effective for coming up with numerous critical keywords.

Start by generating a list of terms and phrases that are relevant to your industry and pertain to what your site or business offers. The brainstorming phase should ideally result in a list of

several dozen to several hundred or more keyword searches that will bring relevant, qualified visitors to your site.

One easy way to begin this process is to gather your team in a conference room and then follow these steps:

1. Produce a list of key one- to three-word phrases that describe your products/services.
2. Spend some time coming up with synonyms that your potential customers might use for those products and services. Use a thesaurus to help you with this process.
3. Create a taxonomy of all the areas of focus in your industry. It can be helpful to imagine creating a directory for all the people, projects, ideas, and companies connected to your site. You can also look at sites that are leaders in the industry and study their site hierarchy as a way to start your thinking about a taxonomy.
4. Broaden your list by thinking of higher-level terms of which your products or services are a subset.
5. Review your existing site, and extract what appear to be key phrases from your site.
6. Review industry association and/or media sites to see what phrases they use to discuss your topic area.
7. List all your various brand terms.
8. List all your products. If your site has a massive number of products, consider stepping back a level (or two) and listing the categories and subcategories.
9. Have your team step back and imagine they are a potential customer, and ask them what they would type into a search engine if they were looking for something similar to your product or service.
10. Supplement this by asking some people outside your business what they would search for, preferably people who are not directly associated with the company.
11. Use your web analytics tool to see what terms people are already using to come to your site.

Gathering this type of intelligence is what a traditional marketer might have done prior to initiating a marketing campaign before the Web existed. And of course, if any of this data is available to you from other departments of the company, be sure to incorporate it into your research process.

Include Competitive Analysis

Your competitors face the same problem, and unless you are very lucky, they are also probably resourceful and creative. You can likely count on their having invested in learning how their customers think and the best ways to appeal to them. So, add these steps to the process:

1. Review your competitors' websites and see what key phrases they use for their products and services that compete with yours.

2. Record what nonbrand terms they use for their business.
3. Read any articles they have written that are published on sites other than their own.
4. Observe what the media may have had to say about them.

Add these ideas into the mix and you will have a wonderfully robust set of keywords to use as a starting point.

You may ask why you should go through all this trouble. Don't the keyword tools take care of all this for you? There are two reasons why the extra effort is critical:

- Your internal team has a rich array of knowledge that the keyword tools do not: they know where to start. Keyword tools require the initial input of information, and the quality of the data they provide is only as good as the quality of the "seeds" you give them.
- The upfront brainstorming helps your organization's stakeholders better understand the market and the opportunities.

Once you have completed these steps you will have in hand a rich set of terms of interest. The next step is to expand those terms of interest using keyword research tools.

Keyword Research Tools

A wide variety of options are available for performing keyword research, including tools provided by the search engines, tools developed by third parties, and tools for complex keyword analysis of terms culled during research. In this section, we will review each of these; but first we'll examine what you can determine just by using the search engines themselves.

Keyword Research Data from the Engines

The search engines provide a number of tools that can help you with keyword research. Many of these are not designed specifically for that purpose, but they can be used to obtain interesting keyword research information if they are used in the right manner.

Blog search counts

Blog search data is terrific for picking out hot topics or keywords in the blogosphere and the realm of social media. Since blog search often incorporates forums and other social media properties (anyone with a feed, really), it is a great way to see how a term/phrase is looking in the social space. Be aware, though, that the data is temporal—anything that's more than a few months old is likely to be out of the blog index (this does make the data a great temporal tracking tool, however). For example, check out the 239,818 results returned by the blog search for *cupcake recipes* (see [Figure 5-2](#)) versus the 1.89 million results returned when web search was used to perform the same search.



FIGURE 5-2. Google blog search counts

Related terms

Several of the engines offer “related” terms, including Google, Yahoo!, Bing, Ask, and Clusty (which shows related terms in clusters, as shown in Figure 5-3). This data can be invaluable if you’re looking to find related terms that may not have come up through competitive analysis or brainstorming.

The screenshot shows the Clusty search interface. The search term 'whale' is entered in the search bar. The results page has tabs for 'clusters', 'sources', and 'sites', with 'clusters' selected. It displays 'Top 257 results of at least 15,500,000 retrieve'. On the left, a sidebar lists 'All Results (261)' with clusters like '+ Whale Watching (54)', '+ Pictures (54)', '+ Research (27)', '+ Whales, Dolphins (24)', '+ Humpback Whale (26)', '+ Cetaceans (11)', '+ Facts (13)', '+ Blue Whale (12)', '+ Definition (6)', and '+ Songs (7)'. Below this is a 'find in clusters:' input field and a 'Find' button. The main content area shows a list of 5 results:

- [See Life As Animals Do.](#) - National Ge
- ["Shark Week" on Discovery](#) - The Mc
- [Whale - Wikipedia, the free enc](#)
en.wikipedia.org/wiki/Whale - [cache] - Li
- [Go to WhaleNet](#) 🔍🔍🔍
Educational site that focuses on **whales**
whale.wheelock.edu - [cache] - Live, Ope
- [WDCS - Whale and Dolphin Cor](#)
Provides information about how you can h
and **whale** watching.
www.wdcs.org - [cache] - Gigablast, Ope
- [Welcome to The Whale Museum](#)
The **Whale** Museum promotes stewardsh
www.whalemuseum.com - [cache] - Live,
- [Whale Center of New England -](#)

FIGURE 5-3. Clusty related terms clusters

Common usage and phrase combinations

Using a search with the * character can give you a good idea of what terms/phrases commonly precede or follow a given term/phrase. For example, using * *ringtones* can show you phrases that are commonly associated with the term *ringtones*, as shown in [Figure 5-4](#).

The screenshot shows a Google search results page. The search bar contains the query "* ringtones". Below the search bar, there are "Search" and "Advanced Search" buttons, along with "Preferences". A "Web" filter is selected. The results are as follows:

- 3gforfree.com | Ringtones and Screensavers for your mobile phone**
Create and upload your own screensavers; Create and upload your **own ringtones**; Share content with other 3gforfree members; Access your content from your ...
www.3gforfree.com/index.php - [Cached](#) - [Similar](#) -
- Mating call Ringtones - eNature | Ringtones**
Download free wildlife and **nature ringtones** to your phone -- no gimmicks or hidden charges.
wild.enature.com/ringtones/ - [Cached](#) - [Similar](#) -
- Myxer - Download Ringtones and More**
Free Ringtones by Myxer is the easiest way to MAKE, SHARE and DOWNLOAD **Free Ringtones**.
www.myxer.com/ - [Cached](#) - [Similar](#) -

FIGURE 5-4. Finding common phrases

Frequency of recent usage

Using the very cool Google date range operator, shown in [Figure 5-5](#), you can determine how many times in the past day, week, month, or year your term has appeared. There are two ways to do this. One is to go into Advanced Search mode within Google and click the Date drop-down box. Once you do that, you can pick from “anytime” (which is the default), “past 24 hours,” “past week,” “past month,” and “past year.” This will limit you to the results that were added to the index during the referenced time frame.

The screenshot shows a Google search results page. The search bar contains the query "blogrush". Below the search bar, there are "Search" and "Advanced Search" buttons, along with "Preferences". A "Web" filter is selected, and "Past 24 hours" is chosen from the dropdown menu. The results are as follows:

Results 1 - 10 of about 689

- Blogrush | supreme-host.com**
4 hours ago - And again your content or ads will appear tool Great, right? But there's more... You can use real-time desktop alerts to inform users with breaking news, ...
supreme-host.com/tag/blogrush - [Similar](#)
- Rush Blog - Rush is a Band Blog: DRUM! Magazine's Drummie awards ...**
14 hours ago - Rush is a Band: Neil Peart, Geddy Lee, Alex Lifeson.
www.rushisaband.com/display.php?id=1900 - [Cached](#) - [Similar](#)

FIGURE 5-5. Google pages indexed in past 24 hours

For additional flexibility, you can perform a normal search, get your result, and add a parameter to the end of the results page URL, using the operators shown in [Table 5-1](#).

TABLE 5-1. Google date search operators

&as_qdr=d	Past 24 hours
&as_qdr=d4	Past four days
&as_qdr=w	Past week
&as_qdr=w5	Past five weeks
&as_qdr=m6	Past six months
&as_qdr=y2	Past two years

This can give you some seasonal data if you follow it closely, and it can also show you who is producing content in your arena. For example, try a search for *blogrush* (*past 24 hours*). You can also get information on activity level from Yahoo! News, as shown in [Figure 5-6](#).



FIGURE 5-6. Yahoo! News activity level

Both Google News and Yahoo! News are great places to do a bit of digging into anyone who is publishing press releases or getting news coverage on the terms/phrases you might be researching. If there's a lot of activity in these arenas (and it is not all PR spam), you can bet the terms are going to be even more competitive. For example, the following URLs Michael Jackson show SEOMoz at Google News (http://news.google.com/news?hl=en&lr=&q=seomoz&as_qdr=d&um=1&ie=UTF-8&sa=N&tab=wn) and at Yahoo! News (<http://news.search.yahoo.com/news/search?p=seomoz&c=>).

You can combine all of this data to form a very well-rounded view of a particular term or phrase, and although it is probably overkill for most keyword research projects, it is certainly a valuable exercise and something to monitor closely if you're basing a lot of your success off of a single search query (or just a handful of queries). Even if you're just trying to get a better sense of what's going on infrequently and informally, these pieces of the keyword puzzle can be remarkably valuable.

Keyword Research with Tools

It is great to get this data from search engine queries, and it can certainly help you get a sense of the importance of a given keyword. However, a large array of tools exists to give you direct

insight into the volume of searches performed on specific keywords, and also to help you discover new keywords to consider.

We review many of these leading tools on the pages that follow.

Google's AdWords Keyword Tool and Traffic Estimator

Google provides a couple of tools specifically designed for use in keyword research. Although they are primarily meant to help their paid search customers, they can also be used to obtain information for organic search.

What the Keyword Tool provides. [Google's AdWords Keyword Tool](#) provides related terms, search volume estimates, search trends, and ad cost estimates for any keyword or URL that you enter (see [Figure 5-7](#)).

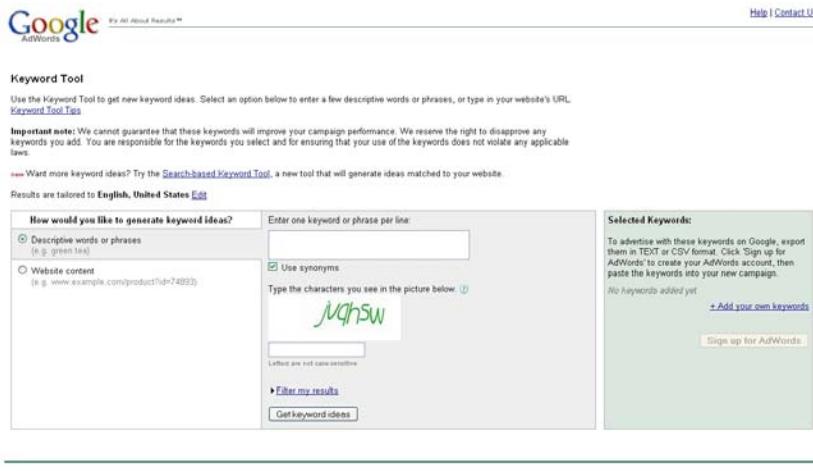


FIGURE 5-7. Google's AdWords Keyword Tool

The AdWords Keyword Tool provides two tabs: Keyword Variations and Site-Related Keywords. From the Keyword Variations tab, you can enter a keyword and the AdWords Keyword Tool will return keywords related to the term you entered and the match type (where you can specify whether you want your search-targeted keywords to be a broad, exact, or phrase match).

You can also choose to display the following information about each keyword. Set the Match Type to Exact to get a sense of the search volume associated with each keyword, or set it to Broad to get a sense of the volume associated with the keyword and the long tail variations of the keyword.

Keyword Search Volume

Displays the keyword, related terms, search volume from the past month, advertiser competition, and match type (see [Figure 5-8](#)).

Cost and Ad Position Estimates

Displays each keyword's estimated average cost-per-click and estimated ad position.

Search Volume Trends

Displays each keyword's average search volume, search volume trends over the course of one year, and in which month the highest search volume occurred (see [Figure 5-9](#)).

Possible Negative Keywords

Allows you to add a negative keyword for any keyword phrase that does not pertain to your business. This feature is not necessarily useful for researching keywords for organic search; rather, it is more valuable when planning your AdWords account bids.

You can also opt to have the results for your keyword show synonyms for your term, which is a great way to find similar, relevant terms for your keywords.

The screenshot shows the Google AdWords Keyword Tool interface. At the top, there's a navigation bar with links like 'All About Results' and 'Help | Contact Us'. The main section is titled 'Traffic Estimator' with a sub-instruction: 'Get quick traffic estimates for new keywords without adding them to an account or using the AdWords sign-up wizard.' Below this, there are four numbered steps:

- 1. Enter keywords, one per line:** A text input field with placeholder text 'Enter individual CPCs using this format: keyword + \$0.25' and a note about optional individual CPCs.
- 2. Choose a currency.** A dropdown menu set to 'US Dollars (USD)' with a note: 'Suggested value should deliver ad in the top position 85% of the time.'
- 3. Choose daily budget (optional).** A text input field for specifying the amount to spend daily.
- 4. Select targeting.** This section has three parts:
 - a. Language:** A dropdown menu showing 'English' as the selected language, with other options like Chinese (Simplified), Chinese (Traditional), Danish, Dutch, Finnish, French, German, Italian, Portuguese, Spanish, Swedish, and Turkish.
 - b. Location Targeting:** Radio buttons for 'Countries and territories - choose countries', 'Regions and cities - choose states and regions and/or enter cities', and 'Customized - enter a radius and address or coordinates'. The 'Countries and territories' option is selected.
 - c. Countries:** A list of 'Available Countries and Territories' including Australia, Austria, Belgium, Brazil, Canada, France, Germany, Italy, Japan, Mexico, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and United States. An 'Add >' button is next to the list, and a 'Selected Countries and/or Territories' sidebar shows 'All Countries and Territories'.

At the bottom, there's a 'Continue >' button and a footer with links: '©2009 Google - AdWords Home - Advertising Policies - Privacy Policy - Contact Us'.

FIGURE 5-8. AdWords Keyword Tool Keyword Variations tab

The screenshot shows the Google AdWords Keyword Tool interface. At the top, there's a search bar with placeholder text "Enter one keyword or phrase per line:" followed by a text input field containing "search engine optimization". Below the search bar are two radio button options: "Descriptive words or phrases (e.g. green tea)" and "Website content (e.g. www.example.com/product?ID=76002)". A checkbox labeled "Use synonyms" is checked, with a link "Filter my results" next to it. A button "Get keyword ideas" is at the bottom of this section.

To the right, a sidebar titled "Selected Keywords" contains a note about advertising with these keywords on Google, exporting them, and creating an AdWords account. It also says "No keywords added yet" and has a link "+ Add your own keywords" and a "Sign up for AdWords" button.

The main content area displays a table titled "Keywords related to [term] entered - sorted by relevance". The columns are: Keywords, Advertiser Competition (with a bar chart icon), Local Search Volume: June (with a bar chart icon), Global Monthly Search Volume (with a bar chart icon), and Match Type: Broad (with a dropdown arrow). The table lists various SEO-related terms with their respective search volumes and competition levels.

Keywords	Advertiser Competition	Local Search Volume: June	Global Monthly Search Volume	Match Type: Broad
search engine optimization tools	4,400	5,400	Add	
search engine optimization tips	22,200	6,600	Add	
search engine optimization seo	60,500	33,100	Add	
search engine optimization services	39,100	40,500	Add	
search engine optimization	49,500	165,000	Add	
google search engine optimization	12,100	6,600	Add	
organic search engine optimization	6,600	8,100	Add	
top search engine optimization	5,400	2,400	Add	
search engine optimization techniques	1,900	2,900	Add	
search engine optimization articles	1,000	880	Add	
search engine optimization consultants	1,900	2,900	Add	
search engine optimization consulting	6,600	5,400	Add	

FIGURE 5-9. AdWords Keyword Tool Search Volume Trends data

By choosing the Site-Related Keywords tab, you can enter a web page URL and AdWords will return various keywords grouped by like terms (see [Figure 5-10](#)).

Like the Keyword Variations option, you can opt to display the following information about each URL's keywords:

- Keyword search volume
- Cost and ad position estimates
- Search volume trends

You can also opt to have AdWords display keyword suggestions for other pages on your site that are linked from the URL you provided, and you can choose to ungroup keywords by common terms.

What the Traffic Estimator provides. Within Google AdWords is a tool called the Traffic Estimator (see [Figure 5-11](#)) that allows you to get estimates of traffic on different keywords (i.e., the potential click-throughs you may see to your site, instead of just the number of impressions provided by tools such as Google's AdWords Keyword Tool). Although this tool is built into AdWords, it can provide invaluable information regarding the traffic data Google has available, which can be useful in organic search optimization. You can also see the tool at <https://adwords.google.com/select/TrafficEstimatorSandbox>.

Keyword Tool

Use the Keyword Tool to get new keyword ideas. Select an option below to enter a few descriptive words or phrases, or type in your website's URL.

[Keyword Tool Tips](#)

Important note: We cannot guarantee that these keywords will improve your campaign performance. We reserve the right to disapprove any keywords you add. You are responsible for the keywords you select and for ensuring that your use of the keywords does not violate any applicable laws.

Want more keyword ideas? Try the [Search-based Keyword Tool](#), a new tool that will generate ideas matched to your website.

Results are tailored to **English, United States** [Edit](#)

How would you like to generate keyword ideas?		Enter one keyword or phrase per line:				
<input checked="" type="radio"/> Descriptive words or phrases (e.g. green tea)		<input type="text" value="search engine optimization"/> <input checked="" type="checkbox"/> Use synonyms <input type="button" value="Filter my results"/> <input type="button" value="Get keyword ideas"/>				
<input type="radio"/> Website content (e.g. www.example.com/product?id=74893)						

Choose columns to display: ? Show/hide columns						
Keywords	Advertiser Competition ?	Local Search Volume: June ?	Global Monthly Search Volume ?	Search Volume Trends (Jul 2008 - Jun 2009) ?	Match Type: ?	Broad ▼
Keywords related to term(s) entered - sorted by relevance ?						
search engine optimization tools		4,400	5,400		Add	
search engine optimization tips		22,200	6,600		Add	
search engine optimization seo		60,500	33,100		Add	
search engine optimization services		33,100	40,500		Add	
search engine optimisation		49,500	165,000		Add	
google search engine optimization		12,100	6,600		Add	
organic search engine optimization		6,600	8,100		Add	
top search engine optimization		5,400	2,400		Add	

FIGURE 5-10. AdWords Keyword Tool Site-Related Keywords tab

When you enter one or more keywords in the Traffic Estimator, the tool will return estimates of the search volume for each term, their average cost-per-click, their ad positions, the clicks per day, and the cost per day. The cost information can provide you with additional insight into how competitive a keyword is in organic search as well.

The Traffic Estimator is located under the Tools section in Google AdWords. Enter your keyword in the box (enter one keyword if you want to see data for just that keyword; enter more keywords if you want a comparison).

You can enter your keyword in the following ways:

Broad match

Entering your keyword without any parameters means it will be broadly matched; this means if you buy an ad for this keyword, it will appear in the search results when any of the words in your keyword phrase are combined and used in a search. For example, your ad for search engine optimization will appear in the results for a search on *search for train engine optimization*.

The screenshot shows the Google AdWords Keyword Tool interface. At the top, there's a search bar with the URL 'http://elchemistmedia.com'. Below it, there are two radio button options: 'Descriptive words or phrases (e.g. green tea)' and 'Website content (e.g. www.example.com/product?id=74893)', with the second one selected. To the right of the URL input, there's a checkbox for 'Include other pages on my site linked from this URL'. Below the URL input, there's a text field for 'Or, enter your own text in the box below (optional)' and a 'Filter my results' button. A large 'Get keyword ideas' button is at the bottom. The main results area has a header 'Choose columns to display' with a dropdown menu and a checked checkbox for 'Group keywords by common terms'. Below this, a sub-header says 'Showing keywords grouped by these terms: search engine optimization (37), pay per click (24), search engine marketing (14), search engine ranking (12), search engine (18), internet marketing (11), search (12), optimization (8), seo (35), marketing (10), services (5), Miscellaneous keywords (14)'. The main table lists keywords under 'Keywords' with columns for 'Advertiser Competition', 'Local Search Volume: June', 'Global Monthly Search Volume', and 'Match Type: Broad'. The table includes rows for various SEO-related terms like 'search engine optimization service', 'affordable search engine optimization', etc.

FIGURE 5-11. Google's Traffic Estimator

Exact match

Putting brackets around your keyword (e.g., *[search engine optimization]*) means your ad will show only when a user types in the exact keyword phrase you are targeting.

Phrase match

Adding quotation marks around your keyword (e.g., “*search engine optimization*”) means your ad will show when a user types in a phrase that contains your keyword. For example, your ad will show on a search for “*how to do search engine optimization*”.

Negative match

Using the minus sign/dash in front of an undesired keyword (e.g., *-spam*) before your keyword (e.g., “*search engine optimization*” for a phrase match) means the term does not apply to your services; therefore, your ad won’t show for the corresponding search (e.g., “*search engine optimization spam*”).

When using the Traffic Estimator for keyword research it is best to enter your keywords as “exact match” for direct comparison.

After you’ve entered your keywords, you can leave Daily Budget blank. Set the currency and then select your language and the location you’re targeting (for U.S.-focused campaigns, use the default of “Countries and territories” and enter “United States”). See [Figure 5-12](#).

Traffic Estimator
Get quick traffic estimates for new keywords without adding them to an account or using the AdWords sign-up wizard.

- Enter keywords**, one per line:
search engine optimization

keyword = broad match
"keyword" = match exact phrase
[keyword] = match exact term only
-keyword = don't match this term

Set optional individual CPCs using this format:
Keyword ** 0.25

- Choose a currency**. Enter a specific Max CPC for your estimates, or leave the field blank and we'll provide estimates based on our suggested Max CPC.*
US Dollars (USD \$)

*Suggested value should deliver ads in the top position 85% of the time.

- Choose daily budget (optional)**. See estimates for the amount you want to spend each day, or leave field blank to see all available clicks.

- Select targeting.**
 - Language**
English
Chinese (simplified)
Chinese (traditional)
Danish
Dutch
Finnish
French
 - Location Targeting**
 Countries and territories - choose countries
 Regions and cities - choose states and regions and/or enter cities
 Customized - enter a radius and address or coordinates

c. Countries

Available Countries and Territories

All Countries and Territories
Australia
Austria
Belgium
Brazil
Canada

Selected Countries and/or Territories

United States

FIGURE 5-12. Traffic Estimator setup screen

When you click Continue, you’ll see data for each keyword you entered. Useful data for keyword research purposes includes Estimated Clicks/Day and Cost/Day. You can compare each keyword’s estimated clicks to see which term is more likely to be searched for and clicked on than others.

In the results shown in [Figure 5-13](#), *internet marketing* is estimated to have 36 to 45 clicks per day, while *search engine marketing* has 10 to 12, *search engine optimization* has 13 to 16, and *seo*

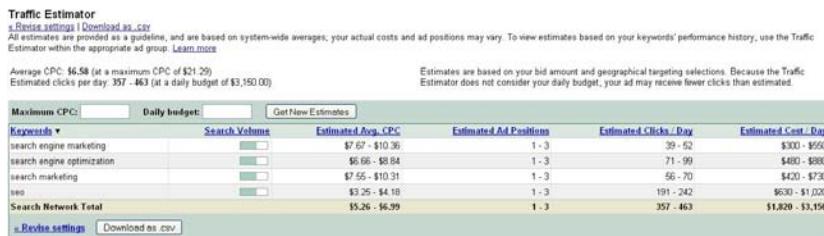


FIGURE 5-13. *Traffic Estimator output*

has 15. Based on this data, it is clear that *internet marketing* is the most popular term of the four and is likely to be one of the more competitive terms. In this particular case, an additional factor enters into the equation, because *seo* is a “trophy term” on which people put an extra focus for branding reasons. Nonetheless, the value of the traffic data is considerable.

Where the tools get their data. Google’s Keyword Tool Estimator and Traffic Estimator get their data from Google’s search query database.

How the tools are useful. The AdWords Keyword Tool offers a lot of useful information about your keyword campaigns, such as suggestions for similar keywords, an estimate of the keyword’s popularity, ad costs and positions, general search volume trend information, and keyword campaign suggestions for your site or your competitor’s site. The tool is great for compiling a lot of general information about a keyword.

The Traffic Estimator provides a decent estimate of your keyword’s click-through rate. Based on the estimated clicks per day, you can get a relative idea of which of your keywords are the most popular and can potentially bring you the most traffic.

Cost. The Keyword Tool is free to use. The Traffic Estimator is also free, but unlike the Keyword Tool, it is not accessible from an external URL; you must sign up for an AdWords account to use the Traffic Estimator tool (it costs a minimum of \$5 to activate an AdWords account).

Yahoo! Search Marketing

Yahoo! also provides a tool to help find keywords for paid search campaigns: Yahoo! Search Marketing. As with the Google tools, Yahoo! Search Marketing offers a lot of useful data for your organic search optimization campaigns.

After you select your time zone and geotargeting preference and provide three keywords/phrases that you want to target, the Search Marketing tool will generate a list of related keywords and an Estimated Searches bar. You need to be a Yahoo! Search Marketing customer to use this tool. When you set up an account the tool is provided in step 2 of the sign-up process. Figure 5-14 shows the Yahoo! Search Marketing tool.

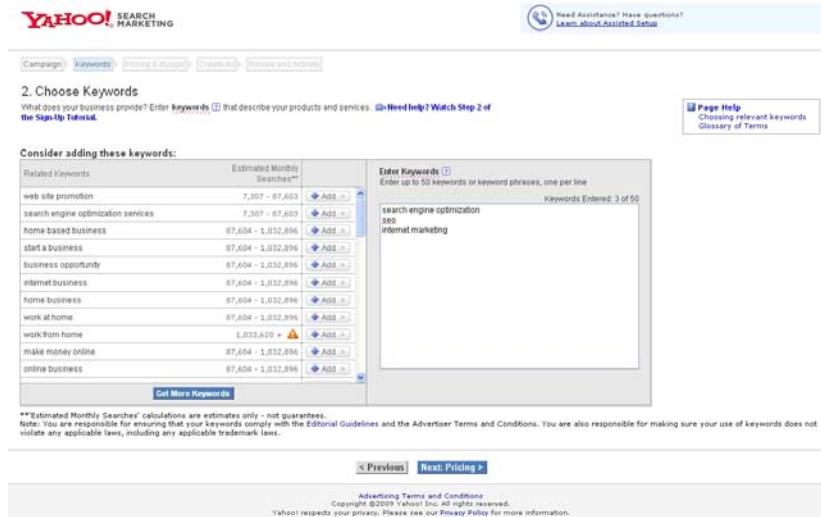


FIGURE 5-14. Yahoo! Search Marketing keyword tool screen shot

Where it gets its data. Yahoo! Search Marketing gets its data from Yahoo!'s search query database. Note that the tool provides no actual search numbers—its gauge of “Estimated Searches” is simply a bar that you can compare with other keywords’ bars for relative popularity.

How it is useful. Step 2 of the Search Marketing sign-up process provides a helpful list of keywords related to the ones you provided. This list is useful in brainstorming various relevant keywords (it can generate terms you had not thought of), and the Estimated Searches gauge can give you a relative idea of which terms are more popular than others.

Cost. A Yahoo! Search Marketing Self-Serve sign-up is free, whereas Assisted Setup costs \$199. Each account requires a minimum \$5 deposit if you plan to advertise on the Yahoo! network.

Microsoft's adCenter Keyword Generation Tool

Microsoft's adCenter Keyword Generation Tool generates keyword suggestions based on a search term or website you enter.

Entering a keyword in the search bar will return data that is divided into two tabs: Contains Term and Similar Terms. The Contains Term tab includes search phrases that contain the keyword you provided, along with the preceding month and the current month of search data figures. For example, a search for *scooter* returns electric scooters, razor scooters, and mopeds. As you can see in [Figure 5-15](#), the term *scooter* had, according to Microsoft, 7,723 searches in the month prior to this screen shot.

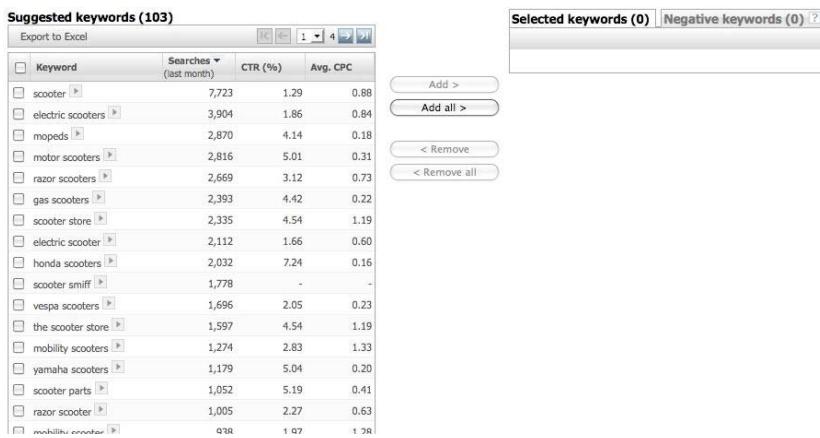


FIGURE 5-15. Microsoft adCenter Keyword Generation Tool basic output

The Export to Excel option allows you to pull the collected data into a spreadsheet. Although the CTR% (click-through rate) and CPC (cost-per-click) columns are intended for paid search customers, they can also provide some indication of SEO value. You can multiply the CTR times the search volume to get a sense of how many clicks a high-ranking paid search result might get (comparable organic results will get six to eight times more clicks), and the CPC provides some indication of the competition for ranking on the term. You can also obtain demographic data as shown in Figure 5-16.

If you enter a URL into the search bar, the tool will return data that, once again, is divided into Contains Term and Similar Terms suggestions.

Where it gets its data. The adCenter Keyword Generation Tool obtains its data from Microsoft's Bing search query database.

How it is useful. This tool is useful in generating keyword suggestions based on a keyword you are targeting or on your site's URL. You can also enter your competitor's URL and see what the keyword suggestions are for their site.

Cost. The adCenter Keyword Generation Tool is free, although you do have to create an account with Microsoft adCenter and provide credit card information in the event that you advertise on the Microsoft network.

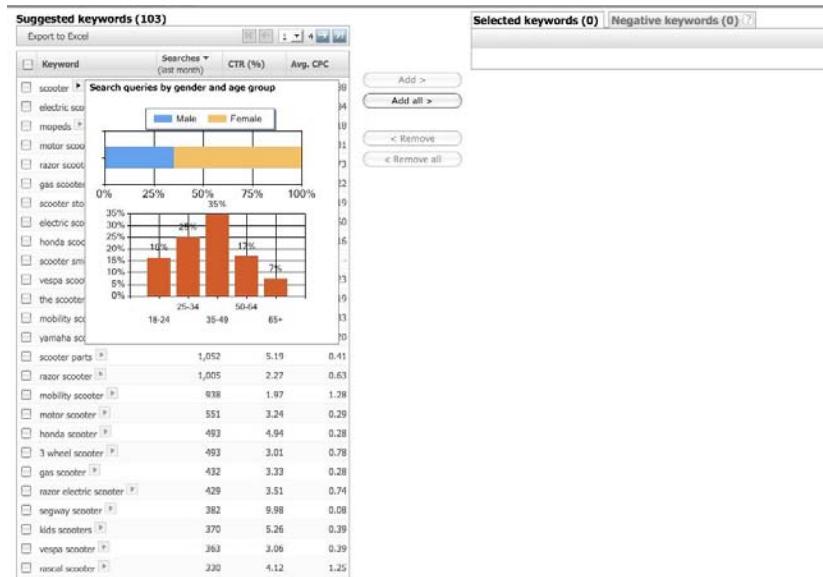


FIGURE 5-16. Microsoft adCenter Keyword Generation Tool demographic data

Wordtracker

Wordtracker is one of the better known keyword tools available that the search engines themselves do not provide. Wordtracker offers the following features:

Keyword Researcher

When you enter a keyword or phrase in the search box under the Research section, Wordtracker displays the most popular search terms that include the keywords you provided, the number of searches per day recorded for those terms, and a prediction of how many searches are typically performed per day (see Figure 5-17).

The Keyword Researcher feature also allows you to evaluate competing search listings with the Evaluate tab, which allows you to analyze competition data from Google, Microsoft, Yahoo!, or other smaller search engines, directories, or news outlets, as well as pay-per-click data from Yahoo!. You get this data by selecting the search engines you want to compare and then clicking the Research button, as shown in Figure 5-18.

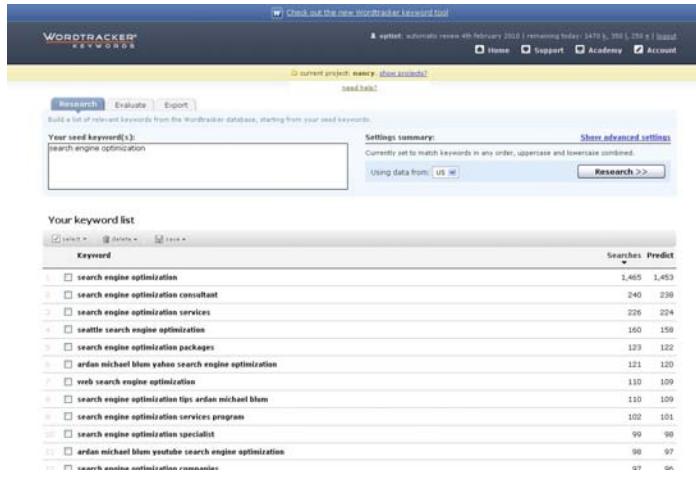


FIGURE 5-17. Sample Wordtracker output

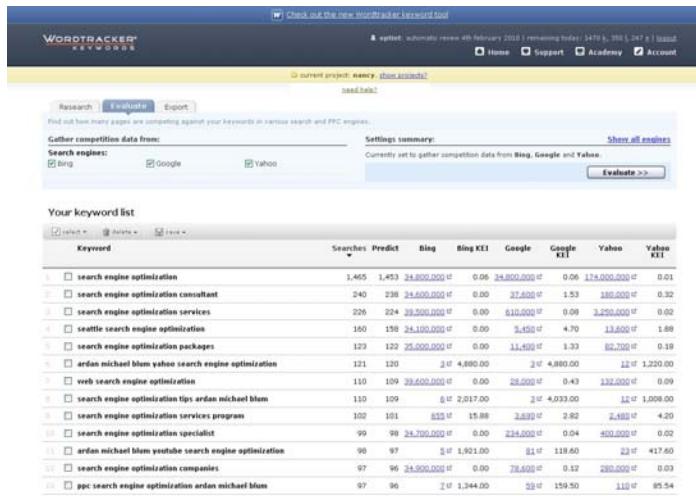


FIGURE 5-18. Wordtracker search engine comparison

Keyword Universe

Keyword Universe provides a list of keywords and phrases that are related to the search term you provide, then gives you a search count for those terms in its Popularity Search. This differs from the other tools in that it can produce terms that are not part of the original seed term you provide. So, if you enter a term such as *search engine optimization* you may get back *internet marketing* in the output.

Full Search

Based on the single generic term you provide, Full Search will return a list of similar keywords and a count of the number of times those keywords appeared in the meta tags of 200 related web pages.

Keyword Projects

Keyword Projects stores your keyword research projects. At any given time, you are allowed one active project and four stored projects (see [Figure 5-19](#)).

Reports

The Short-Term Report provides the top 1,000 search terms from the preceding 48 hours. The Long-Term Report provides the top 1,000 search terms that have been repeatedly or consistently searched for over the preceding 90 days (we declined to include a screenshot of these reports, as many of the terms included in them are a bit risqué and may be offensive to some people).

Free keyword suggestion tool

Wordtracker also has a free keyword suggestion tool (<http://freekeywords.wordtracker.com/>).

When you enter a keyword/phrase, you'll see Wordtracker's count of the total number of searches performed across the Web in the preceding 90 days. You will also see a list containing both the keyword you searched for and similar keywords, along with their predicted daily search count.

Wordtracker also allows you to tweak the parameters of your search. For example, you can choose to include misspellings, allow wildcards, match only the exact keyword, and more. [Figure 5-20](#) provides a look at the Setting screen.

Current Project

Project Name:	Nancy	Change Project Name
Basket:	Wordtracker	Number of Keywords
	73	Reset Export Import

Stored Projects

To make any project the CURRENT project, simply click on 'make current'.

Name: Project7	Make Current	Name: Protrade	Make Current
Basket:	No. of Keywords	Basket:	No. of Keywords
Wordtracker	None	Wordtracker	305

FIGURE 5-19. Wordtracker stored projects

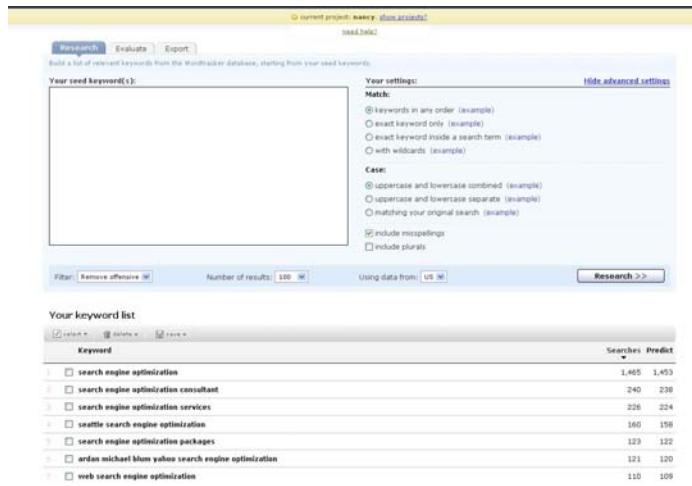


FIGURE 5-20. Wordtracker settings

Where it gets its data. Wordtracker compiles a database of 330+ million search terms from Dogpile and MetaCrawler. This database is updated every week. Nielsen reported Dogpile as having a 0.6% share of total searches across the Web in January 2007, whereas MetaCrawler is estimated to have less than a 0.4% share of total searches.

How it is useful. Wordtracker is great for finding out how many searches were performed on various keywords. Because its data sources are limited, you should not rely on the tool for precise data figures; it is a good tool to use to get a general idea of which keywords are searched for more often than others.

Cost. Wordtracker provides different subscription offerings that range from a one-day membership for \$8.26 (pricing as of October 2009) to a one-year membership for \$275.24 (October 2009 pricing). The free tool with limited features is also available. We recommend checking out the different types and choosing a package that will work best for your company.

KeywordDiscovery

Another popular third-party tool for keyword research is [Trellian's KeywordDiscovery](#). KeywordDiscovery offers the following features:

Keyword Research

When you enter a keyword or phrase in the search bar under the Research section, KeywordDiscovery displays the most popular search terms that include the keywords you provided, along with a count of how many searches were performed for those keywords in the past 12 months (see [Figure 5-21](#)).

Clicking on Analyze brings up more detailed data about the search term. Specifically:

Query

Shows the term you entered as well as related terms.

Searches

Shows the number of times each term was searched for (according to KeywordDiscovery's database) over the past 12 months.

Occurrences

Shows an estimate of the number of pages where each search term appears.

KEI

An acronym for Keyword Effectiveness Indicator, measures the value of a search term. It factors in the number of times a term has been searched for versus how many other web pages target the exact match phrase. The higher the KEI, the more value that term has. For example, *shoes* has a KEI of 1.51 because it is searched for a lot but is also targeted by a lot of web pages. Conversely, *dance shoes* has a KEI of 4.99 because it is searched for less often and is targeted by fewer sites. We don't recommend using the KEI as a metric for search term value because search result numbers do not correlate well to the level of competition for a term.

Predicted Daily

A prediction of how many average daily searches are performed for each keyword across the Web (see [Figure 5-22](#)).

The screenshot shows the KeywordDiscovery software interface. At the top, there is a search bar with 'Search Term: search engine optimization', a dropdown for 'Results per page: 10', a 'Search' button, and a help icon. Below the search bar are several filter options: 'Exclude:' (empty), 'Database: Global Premium' (selected), 'Historical?' (unchecked), and various checkboxes for search parameters like 'Phrase Match?', 'Include Plurals?', 'Remove Spaces?', 'Adult Filter?', 'Spell?', 'Thesaurus?', 'Inflected Form?', 'Competitors?', 'Related?', 'Fuzzy/Like?', 'Industry?', and 'Domain Score?'. Below these filters, the results are displayed. The header for the results table includes columns for 'Estimated results: 211', 'Page: 1 2 3 4 5 6 7 8 9 10 >>', and 'Also Searched Queries - Sor'. The main results table has two columns: 'Query?' (with checkboxes) and 'Searches?' (with numerical values and small icons). The first few rows show: 'search engine optimization' (1,415), 'search engine optimization services' (169), 'search engine optimization tool' (167), 'search engine optimization service' (105), 'search engine optimization certification' (90), 'search engine optimization company' (60), 'dls search engine optimization' (56), and 'search engine optimization tips' (54). To the right of the results table, there is a column titled 'Also Searched Queries - Sor' with a list of terms: 'find it', 'search', 'ron', 'frnepu ratvar bcfgvzvmr', 'ca', 'youtube', 'myspace', and 'socialnetwork'.

Query?	Searches?	Also Searched Queries - Sor
<input type="checkbox"/> Query ?	Searches ?	<input type="checkbox"/> Query ?
<input type="checkbox"/> search engine optimization	1,415	<input type="checkbox"/> find it
<input type="checkbox"/> search engine optimization services	169	<input type="checkbox"/> search
<input type="checkbox"/> search engine optimization tool	167	<input type="checkbox"/> ron
<input type="checkbox"/> search engine optimization service	105	<input type="checkbox"/> frnepu ratvar bcfgvzvmr
<input type="checkbox"/> search engine optimization certification	90	<input type="checkbox"/> ca
<input type="checkbox"/> search engine optimization company	60	<input type="checkbox"/> youtube
<input type="checkbox"/> dls search engine optimization	56	<input type="checkbox"/> myspace
<input type="checkbox"/> search engine optimization tips	54	<input type="checkbox"/> socialnetwork

FIGURE 5-21. KeywordDiscovery basic output

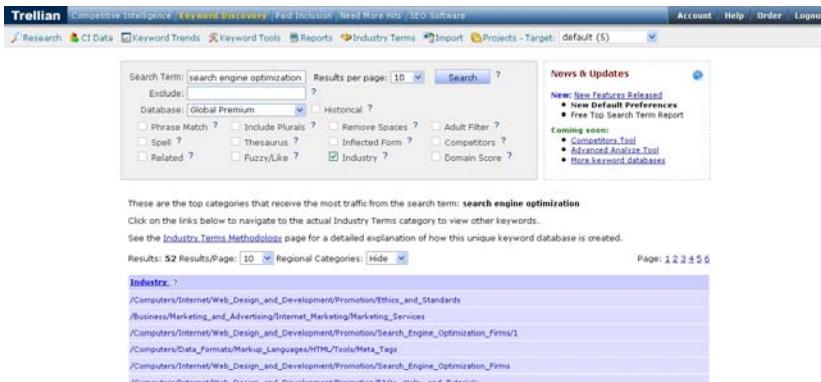
<input type="checkbox"/> Query... ?	Searches... ?	Successful... ?	Results	KCI	PPC	Results	KCI	PPC	Results	KCI	Results	KCI	Daily... ?
<input type="checkbox"/> search engine optimization	1,415  	24.59%	32,000,000	0.10	PPC	 	 	34,800,000	0.09	46,600,000	0.06	363	
<input type="checkbox"/> search engine optimization services	169  	50.30%	537,000	0.73	PPC	 	 	33,700,000	0.01	54,400,000	0.01	43	
<input type="checkbox"/> search engine optimization tool	167  	1.80%	45,800	2.83	PPC	 	 	33,600,000	0.01	7,930	4.71	43	
<input type="checkbox"/> search engine optimization service	105  	79.05%	133,000	1.48	PPC	 	 	39,200,000	0.01	34,100,000	0.01	27	
<input type="checkbox"/> search engine optimization certification	90  	18.09%	4,910	4.62	PPC	 	 	34,700,000	0.01	447,000	0.53	23	
<input type="checkbox"/> search engine optimization company	60  	66.67%	338,000	0.49	PPC	 	 	33,900,000	0.00	49,700	1.83	15	
<input type="checkbox"/> dsl search engine optimization	56  	91.07%	607	6.44	PPC	 	 	115	8.27	64	8.91	14	
<input type="checkbox"/> search engine optimization tips	54  	55.56%	139,000	0.93	PPC	 	 	33,000,000	0.00	17,500,000	0.01	14	
<input type="checkbox"/> search engine optimization companies	35  	94.29%	76,900	1.06	PPC	 	 	33,800,000	0.00	53,900,000	0.00	9	
<input type="checkbox"/> dsl engine optimization search	35  	37.14%	2,550	4.40	PPC	 	 	27,200,000	0.00	280	6.82	9	

FIGURE 5-22. KeywordDiscovery predicted daily output

Industry Keywords

This tool tracks the most popular search terms that are bringing traffic to sites in different industries.

For example, according to KeywordDiscovery, the top 10 search queries related to shopping are *ebay*, *ikea*, *walmart*, *amazon*, *target*, *ebay.com*, *best buy*, *adidas*, *home depot*, and *nike* (see Figure 5-23).



The screenshot shows the Trellian Keyword Discovery interface. In the search bar, 'search engine optimization' is entered. The results page displays a list of categories under 'Industry...?' such as 'Computers/Internet/Web_Design_and_Development/Promotion/Ethics_and_Standards', 'Business/Marketing_and_Advertising/Internet_Marketing/Marketing_Services', etc. A sidebar on the right titled 'News & Updates' lists 'New Features Released', 'New Default Preferences', and 'Top Search Term Report'. It also mentions 'Coming soon:' with links to 'Competitors Tool', 'Industry Research Tool', and 'More Keyword Databases'.

FIGURE 5-23. KeywordDiscovery keywords by industry

Spelling Mistake Research

Typing the query *spell:[keyword]* in the search bar under the Research section will return spelling variations for that keyword, the number of times the keyword has been searched (searches), and the keyword results for your search (queries).

For example, *spell:optimization* returns results such as *search-engine-optimization*, *searchengineoptimization*, and *search-engine-optimisation*, as shown in Figure 5-24.

The screenshot shows the Trellian KeywordDiscovery interface. In the top navigation bar, 'Competitive Intelligence' and 'Keyword Discovery' are selected. The main search term is 'search engine optimization'. The results page displays a table of search terms with their similarity, searches, and queries. A sidebar on the right titled 'News & Updates' includes sections for 'New Features Released' (with links to 'New Default Preferences' and 'Top Search Term Report'), 'Coming soon:' (with links to 'Competitor Tool', 'Advanced Analytics Tool', and 'Market Share databases'), and 'Also Searched Queries - Sorted by user frequency'.

Query	Searches
search engine optimization	1,215,500
find it	2,307,192
search	1,356,777
ron	8,948
fmeputu ratvar bogyzymngvba	346,673
ca	2,217,253
youtube	2,436,595
myspace	63,504
socialnetwork	63,504

FIGURE 5-24. KeywordDiscovery spelling mistakes output

Seasonal Search Trends

If you click on the little bar graph icon next to the number of searches for a query, you'll see a graph of the search trends for that keyword over the past 12 months. You can mouse over each bar and see the number of searches for that time period, and you can sort the chart by historical data (how many searches in the past year), monthly data (number of searches broken down into each month), trends (a graph of the search trends over the past year), combination (a graph of Historical Global versus Global Premium search data; Figure 5-25 shows a definition of these terms), and market share (a breakdown of which search engines were used to search for the query).

The screenshot shows the Trellian KeywordDiscovery interface with the 'Seasonal Search Trends' section active. The search term is 'search engine optimization'. On the left, there's a 'Search Trends' chart showing 'Searches per million users' over a 12-month period. On the right, there's a 'Search Results' table and a 'Search Engine Market Share' section. The 'Search Results' table shows one entry: 'search engine optimization' with 1,415 searches. The 'Search Engine Market Share' section shows 'search engine optimization' with a blue bar indicating its market share.

FIGURE 5-25. KeywordDiscovery seasonal search trends

Related Keywords

Typing either `related:[keyword]` or `crawl:[keyword]` will return keywords that are related to the term you provided. For example, typing in `related:search engine optimization` returns results such as *search engine*, *internet marketing*, *seo*, and *website promotion*.

Keyword Density Analysis

This feature checks how often keywords are found on the URL you provide, assigns a keyword density percentage to those keywords, and lists the number of searches performed for each term.

We do not recommend using keyword density as a metric to judge a page's keyword targeting. The search engines use far more sophisticated analyses of keywords for their algorithms, and relying on rough counts such as this can seriously mislead you. See “[Keyword Targeting](#)” on page 211 in [Chapter 6](#) for more on how to effectively target keywords on the page.

One good use for the Keyword Density Analysis feature is to enter a competitor’s URL into the search bar and see what keywords the site is targeting. It is a great tool to use for competitive research.

Domain Researcher Tool

This tool requires an Enterprise subscription. It allows you to search for available domains that are based on popular keyword search terms. These domains have high traffic potential, as the tool shows how many users have searched for the URL. The tool is great if you want to register other domains in your industry and want these domains to be keyword-rich (see [Figure 5-26](#)).

The screenshot shows a search interface for 'shoes' with results for 'shoes', 'jordan shoes', 'nike shoes', 'dc shoes', 'womens shoes', 'payless shoes', 'shoes store', 'coach shoes', 'vans shoes', and 'black shoes'. To the right, a table lists domain names like 'shoes', 'jordanshoes', 'nikeshoes', etc., categorized by extension (.com and .net), with status indicators for 'REGISTERED' or 'FOR SALE'.

Query ?	Searches ?	Domain Names	.com	.net
<input type="checkbox"/> shoes	124,529	shoes	REGISTERED	REGISTERED
<input type="checkbox"/> jordan shoes	11,917	jordanshoes	REGISTERED	REGISTERED
<input type="checkbox"/> nike shoes	10,488	nikeshoes	REGISTERED	REGISTERED
<input type="checkbox"/> dc shoes	9,560	dcschoes	REGISTERED	REGISTERED
<input type="checkbox"/> womens shoes	7,013	womensshoes	REGISTERED	REGISTERED
<input type="checkbox"/> payless shoes	5,884	paylessshoes	REGISTERED	FOR SALE
<input type="checkbox"/> shoes store	4,929	shoesstore	REGISTERED	REGISTERED
<input type="checkbox"/> coach shoes	4,501	coachshoes	REGISTERED	REGISTERED
<input type="checkbox"/> vans shoes	4,464	vansshoes	REGISTERED	AVAILABLE
<input type="checkbox"/> black shoes	4,113	blackshoes	REGISTERED	REGISTERED

FIGURE 5-26. KeywordDiscovery Domain Research Tool

Competitive Intelligence reports

Trellian, which powers KeywordDiscovery, also offers various Competitive Intelligence reports (which require a separate subscription). These reports include:

Link Intelligence

Identifies which links are sending traffic to your competitors

Search Term Intelligence

Identifies which search terms/phrases are driving traffic to your competitors

Search Engine Intelligence

Identifies which specific search engines send traffic to your competitors

PPC Campaign Intelligence

Identifies which search terms your competitors are bidding on

Referrer Intelligence

Provides information about specific sites that are referring traffic to your competitors

Popularity Index Report

Monitors the Popularity Index (which is based on the number of unique sessions a domain receives) of your competitors

Ranking Report

Provides a view of which terms your competitors are ranking for, the rank of these terms, and any changes in ranking over the past 30 days

Meta keywords

Provides a report that analyzes your competitors' meta keywords (see [Figure 5-27](#))

Competitive Intelligence Executive Report

Provides information about every Competitive Intelligence Report available, as well as several subreports

Free Keyword Suggestion Tool

KeywordDiscovery offers a [free keyword research tool](#) that is similar to Wordtracker's free Keyword Suggestion Tool. When you enter a keyword/phrase, you'll see a list containing both the keyword you searched for and similar keywords, along with their predicted search count over the past 12 months.



FIGURE 5-27. KeywordDiscovery Competitor Meta Keywords Report

Where it gets its data. Trellian derives its keyword data primarily from aggregated Historical Global data purchased from ISPs. Trellian also uses a panel of 4.4 million users to collect its Global Premium data. The company touts that the Global Premium data removes the bias that various spiders introduce into data from other sources.

How it is useful. As we mentioned earlier, KeywordDiscovery offers a multitude of tools that are great for keyword research. Trellian also offers various tools that are useful for competitive

research. You can almost think of KeywordDiscovery as a one-stop shop for research since it offers a diverse set of tools, but as with many of the other keyword research tools we've discussed here its data sources are limited, and you need to take this into account in your use of the tool.

Cost. KeywordDiscovery offers different subscription options that range from a standard monthly subscription for \$69.95 to a yearly Enterprise subscription for \$4,455 (pricing is based on March 2009 prices for the product). Competitive Intelligence Reports range from \$99.95 per month per domain (plus a \$150 setup fee) to \$995 per year per domain. The free tool with limited features is also available. We recommend reviewing the options and choosing the package that will work best for your company.

Google Trends

Google Trends (<http://www.google.com/trends>) allows you to compare two or more search terms to each other to see relative popularity and seasonality/trending over time. If you enter the terms into the search bar and separate them with a comma, you'll see the requested terms' trend history depicted in different colors on a graph spread over a certain time period. You can modify the results by changing the time period and/or region (see Figure 5-28).



FIGURE 5-28. Google Trends sample output

With Google Trends, users can also see Google's estimate of which cities, regions, and languages performed the largest number of searches for a particular keyword (see Figure 5-29).



FIGURE 5-29. Google Trends top cities data

Experienced marketers often feel that this data is imprecise (and occasionally inaccurate) because more accurate data from analytics and search advertising campaigns have often contradicted the results.

Lastly, plotted on each graph are a few articles/search results related to your keyword query, which correlate to peaks and valleys in the historical search popularity.

Where it gets its data. Google Trends gets its data from searches performed on Google.

How it is useful. Google Trends is a great, easy tool for comparing keywords and identifying which is more popular than the other; in addition, you can examine this data over many years with seasonality factored in. Although Google Trends doesn't supply figures, the graphs are simple to understand and provide a perfect visual of search trends over a particular period of time. Note that this works only with relatively popular terms, not with long tail search terms.

Cost. Google Trends is free to use.

Hitwise

Hitwise offers a wide range of competitive and web statistics via its service. One component of the Hitwise suite, Hitwise Search Intelligence, is a powerful keyword research tool for analyzing the long tail of search data. It provides extensive insights into how people have successfully searched for products and services across all major search engines, including the breakdown of paid and organic traffic (you can read more about the Hitwise product offering in "[Tying SEO to Conversion and ROI](#)" on page 388 in Chapter 9).

Hitwise Search Intelligence provides the following features:

- Timely information on search terms your specific competitors use
- Market-specific results, for taking advantage of cultural differences on how people search locally
- Information on terms that have been "clicked on" before visiting a website or industry

Figure 5-30 shows an example of the most popular search terms used prior to visiting eBay.

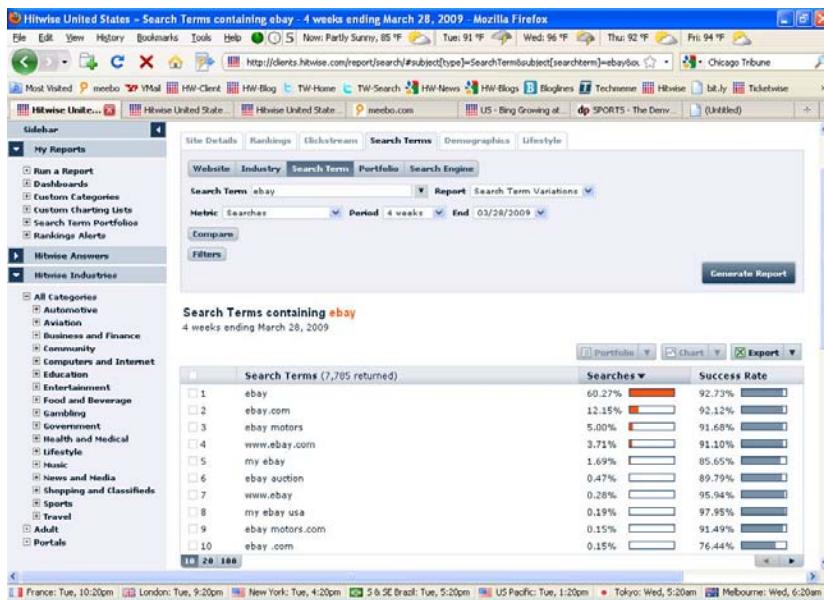


FIGURE 5-30. Hitwise “popular search terms” report

The ability to see actual keyword data on your competitors is an extremely potent feature. You can see what is working for them and what is not. This type of information is very powerful and can give you a significant edge over the competition.

You can also focus more directly on search term suggestions, as shown in Figure 5-31, which depicts a screen shot for terms related to *ipod*.

Where it gets its data. Hitwise derives its data from more than 25 million people’s interaction with the Internet (10 million from the United States). Hitwise collects anonymous Internet usage information from a combination of ISP data partnerships and opt-in panels.

How it is useful. The data is presented in percentages (the volume of searches, its success rate with searchers), which makes it very easy to compare the relative popularity of various keywords, but difficult to estimate the actual number of searches for a given term.

Cost. Hitwise is not an inexpensive tool. The website does not list pricing information, but you should be ready to spend \$20,000 if you plan to engage with this tool. Bear in mind that we have presented only a snapshot of its features, and the competitive data is extremely valuable, not just to the SEO team but to all marketing disciplines across your organization.

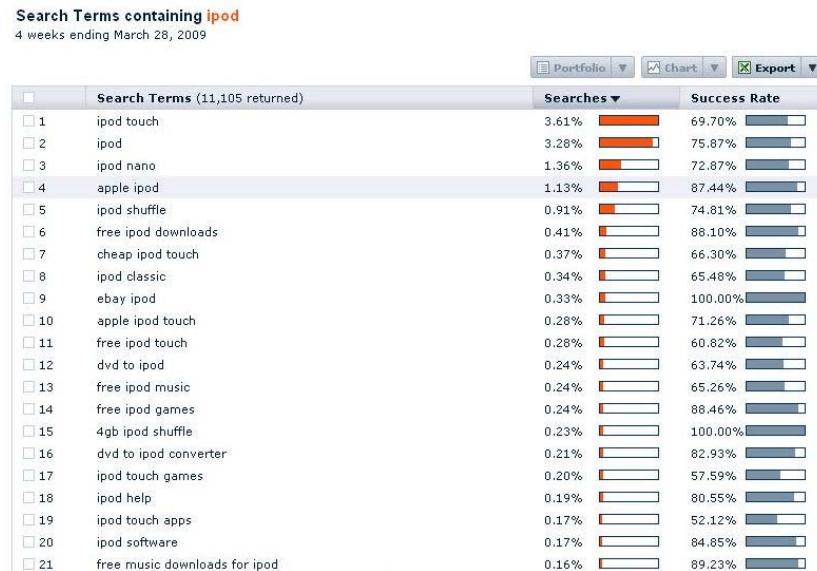


FIGURE 5-31. Hitwise Search Term Suggestions tool

comScore Marketer

Like Hitwise, **comScore Marketer** is a tool that provides a range of data as a result of monitoring the behavior of actual users on the Internet. This data includes details on search terms used, as well as competitive search term analysis. You can read more about comScore Marketer in “[Tying SEO to Conversion and ROI](#)” on page 388 in Chapter 9.

What it provides. ComScore Marketer comprises eight modules, but two of them are particularly useful for keyword research:

Site Profile (for Site(s) X)

This module tells you what search terms and search engines are driving the most traffic to your site, to your competitor’s site, and within your category.

Profile Search Terms

This module tells you the demographic profile of people searching on a set of search terms, as well as what sites these searchers tend to visit.

Figure 5-32 shows the highest-volume terms specific to the Health category.

You can also view similar data specific to a competitor’s site, so you can see what search terms are driving their traffic.

Where it gets its data. ComScore monitors the behavior of approximately 2 million users. These users have voluntarily joined comScore’s research panels in return for free software, free Internet-based storage, or chances to win prizes.

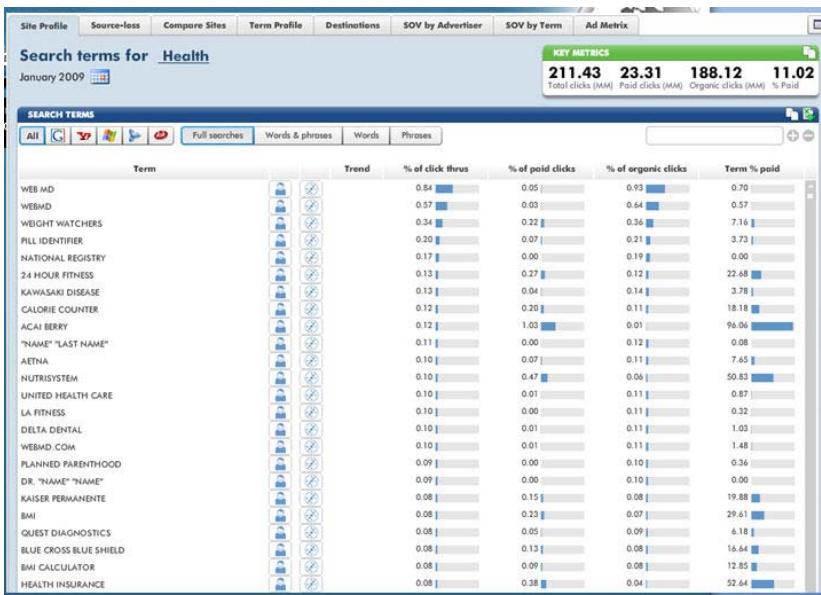


FIGURE 5-32. ComScore “health search terms” report

How it is useful. The data is presented in percentages (the volume of searches, its success rate with searchers), which makes it very easy to compare the relative popularity of various keywords, but difficult to estimate the actual number of searches for a given term.

Cost. Pricing for comScore Marketer is available only upon contacting the company. The primary audience for the product is mid-size to large companies with developed SEM/SEO strategies, but the company has some smaller clients as well.

Enquisite Optimizer

Enquisite Optimizer is a tool that provides some great insights into your long tail search volume. One of its neat features is that it allows you to group terms (e.g., all terms that include the word *seo*) and see how those terms performed as a group. This data from Enquisite Optimizer allows you to analyze not only which long tail terms are bringing in traffic, but also which ones are leading to conversions.

What it provides. Enquisite Optimizer provides a deep and rich look at the long tail of your search volume, as shown in Figure 5-33.

As you can see in Figure 5-33, the first four terms (the “head terms”) make up only a tiny fraction of the overall traffic, while the rest of the terms comprise a considerable majority and yet have brought only one or two visits to the site. The numerical chart in Figure 5-34 shows

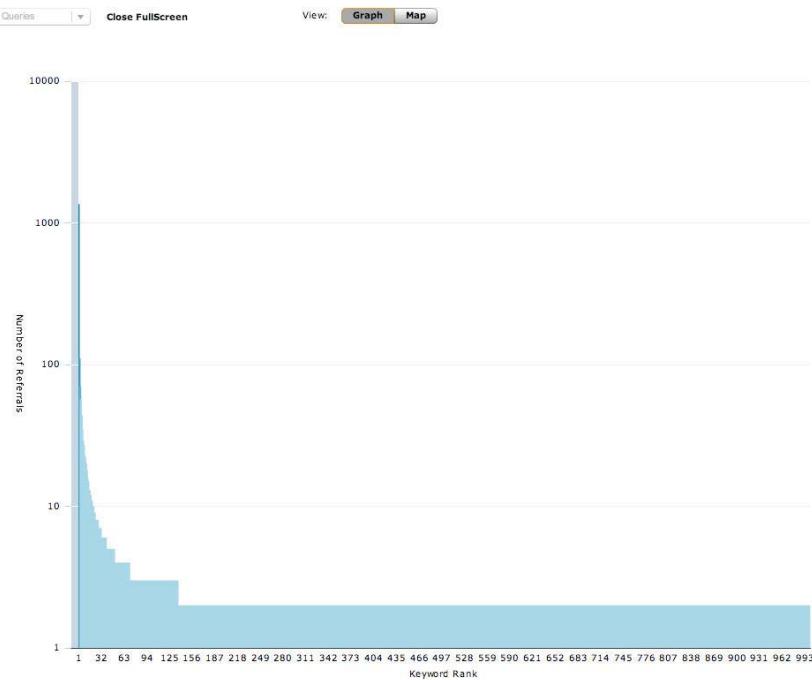


FIGURE 5-33. Enquisite Optimizer sample report

this in greater detail; note that the top phrases bring in only a minuscule percentage of SEOMoz's total search traffic.

Where it gets its data. Publishers incorporate a small amount of JavaScript on their site, and then the tool collects data specific to the site. The JavaScript runs on Akamai's network, so the load time is very minimal.

How it is useful. Because of the data collection method used, the data is specific to the publisher's site. It provides a close look at the long tail terms users are currently using to find your site, and therefore provides you with data you can use to consider ways to expand upon that traffic.

Cost. As of May 2009, low-volume sites (fewer than 100,000 search referrals per month) are priced at \$49.95 per month.

Things to Keep in Mind

It is important to keep in mind that when you are using the various keyword research tools to brainstorm keywords, they are all based on relatively limited data. In addition, each tool will provide different search counts than the other tools. Rather than focusing on the exact search

Graph	View	Phrases	#	%	H	L	A	K
●	eye	marketing	15681	14.48	1	1	1.00	1
○	eye	Netconcepts	6895	6.37	1	1	1.00	1
●	eye	Search	6373	5.89	1	1	1.00	1
●	eye	ecommerce	5466	5.05	1	21	1.02	1
●	eye	articles	3430	3.17	1	1	1.00	1
●	eye	SEO title tag	3111	2.87	1	1	1.00	1
●	eye	sites	2267	2.09	1	1	1.00	1
●	eye	Net Concepts	1958	1.81	1	3	1.00	1

FIGURE 5-34. Enquisite Optimizer sample keyword data

count of various terms, you should think of each tool as a good way to get a general comparison of two search terms.

For example, if you compare two terms and see that one term is more popular than the other because it returns a higher search count, you at least know that Term A is more popular and searched for more often than Term B, but you can treat the search counts as only (rough) estimates.

If you are just starting out with keyword research, consider starting with the Google Keyword Tool and either Wordtracker or KeywordDiscovery. This will give you a rich data set with which to begin your keyword research. Over time you can experiment with the other tools and adjust your process as you find tools that you prefer for one task or another.

Determining Keyword Value/Potential ROI

Once you have the raw keyword data from the research with your favorite tools, you need to analyze which tools have the highest value and the highest ROI. Unfortunately, there are no simple ways to do this, but we will review some of the things you can do in this section.

Estimating Value, Relevance, and Conversion Rates

When researching keywords for your site, it is important to judge each keyword's value, relevance, and potential conversion rate. If a keyword is strong in all three criteria, it is almost certainly a keyword you want to plan to optimize for within your site.

Determining keyword value

When judging the value of a keyword, you should contemplate how useful the term is for your site. How will your site benefit from targeting these keywords?

Identifying relevant keywords

To identify relevant, high-quality keywords, ask yourself the following questions:

How relevant is the term/phrase to the content, services, products, or information on your site?

Terms that are highly relevant will convert better than terms that are ancillary to your content's focus.

Assuming a visitor who searches for that term clicks on your result in the SERPs, what is the likelihood that she'll perform a desired action on your site (make a purchase, subscribe to a newsletter, etc.), create a link to your site, or influence others to visit?

It is a good idea to target keywords that indicate imminent action (e.g., *buy cranium board game*, *best prices for honda civic*), because searchers are more likely to perform the corresponding action on your site when they search for those terms than they are for terms such as *honda civic* or *cranium board game*. Your click-through/conversion rates are likely to be higher if you target keywords that indicate the intent behind the search.

How many people who search for this term will come to your site and leave dissatisfied?

Pay attention to your site's content and compare it to what other sites in the top results are offering—are these sites doing or offering something that you haven't thought of? Do you feel as though these sites offer a more positive user experience? If so, see what you can learn from these sites and possibly emulate on your own.

You can also use an analytics program and check to see which of your pages have the highest abandonment rates. See what you can change on those pages to improve user experience and increase users' level of enjoyment when using your site.

It is important to categorize your keywords into high and low relevance. Generally, keywords of higher relevance will be more beneficial to your site in that they best represent your site as a whole. If, when judging the relevance of a keyword, you answer "yes" to the preceding questions, you've found a highly relevant term and should include it in your targeting.

Keywords with lower relevance than those that lead to conversions can still be great terms to target. A keyword might be relevant to your site's content but have a low relevance to your business model. In this case, if you target that keyword, when a user clicks on your site and finds the content to be valuable she is more likely to return to the site, remember your brand, and potentially link to your site or suggest it to a friend. Low-relevance keywords, therefore, present a great opportunity to strengthen the branding of your site. This type of brand value can lead to return visits by those users when they are more likely to convert.

Determining conversion rates

A common misconception is that a conversion refers only to the purchase of an item on your site. However, many different types of actions users perform can be defined as a conversion, and they are worth tracking and segmenting (you can read more about this in "[Key Performance Indicators for Long Tail SEO](#)" on page 435 in [Chapter 9](#)).

The many different types of conversions create distinct opportunities for targeting various keywords. Although one keyword may work well for purchase conversions, another may be well suited to get users to subscribe to something on your site. Regardless of what type of conversion you are optimizing for, you should strive to have each keyword that you intentionally target convert well, meaning it should be relatively successful at getting searchers to click through to your site and, consequently, perform a specific action.

To know which keywords to target now (and which to pursue later), it is essential to understand the demand for a given term or phrase, as well as the work required to achieve those rankings. If your competitors block the top 10 results and you're just starting out on the Web, the uphill battle for rankings can take months, or even years, of effort, bearing little to no fruit. This is why it is essential to understand keyword competitiveness, or keyword difficulty.

To get a rough idea of the level of competition faced for a particular term or phrase, the following metrics are valuable:

- Search Demand Volume (how many people are searching for this keyword)
- Number of Paid Search Competitors and bid prices to get in the top four positions
- Strength (age, link power, targeting, and relevance) of the Top 10 Results
- Number of Search Results (it can be valuable to use advanced operators such as “exact search” or the `allintitle` and `allinurl` operators here as well; see <http://www.netconcepts.com/google-ebook/> for more on using these specialized searches for keyword research)

SEOMoz offers a [Keyword Difficulty Tool](#) that does a good job collecting all of these metrics and providing a comparative score for any given search term or phrase.

In addition, [Enquisite Campaign](#) provides “optimization difficulty” estimates pertaining to paid search competitiveness that is based on the number of bidders and the cost to be in the top four positions.

Testing Ad Campaign Runs and Third-Party Search Data

One of the things we have emphasized in this chapter is the imprecise nature of the data that keyword tools provide. This is inherent in the fact that the data sources each tool uses are limited. It turns out that there is a way to get much more precise and accurate data—the trick is to make use of Google AdWords.

The idea is to take the keywords you are interested in and implement a simple AdWords campaign. Assuming that you are implementing this campaign solely to get keyword volume data, target position #4 or #5. This should be high enough that your ads run all the time, but low enough that the cost of collecting this data won't be too high.

Once you have run this for a few days, take a look at your AdWords reports, and check out the number of impressions generated for the keyword. Although this data is straight from Google, it is important to remember that the advertisers' ads may not be running all the time, so more (possibly even significantly more) impressions may be available.

Next up, you want to think about the value of achieving certain rankings in the organic results. You can come up with a good estimate of that as well. The key here is to leverage what you know about how click-through rates vary based on organic search position (this is the same AOL data we discussed in “[How People Search](#)” on page 9 in [Chapter 1](#)). **Table 5-2** depicts click-through rates by SERP position.

TABLE 5-2. Click-through rates by SERP position

Organic position	Click-through rate
1	42.1%
2	11.9%
3	8.5%
4	6.1%
5	4.9%

This data, of course, is aggregated across a very large number of searches on AOL, so it serves only as an estimate; but if you are in position #1, the estimate is that 42.1% of the people who will search on a term will click on your result. In the case of our term that is searched 52 times per day, the site in the #1 position will get about 22 clicks per day.

NOTE

There are certain search terms in which these estimates do not apply. For example, if the user searches on a brand term, the focus on the #1 position is much, much higher. Publishers in lower positions still get some traffic, but at lower percentages than we've outlined here.

So, now you have a working estimate of the search volume and the number of clicks per day that the term will deliver. Can you get an estimate of conversion rates as well? Yes, you can. This requires only a simple extension of the AdWords campaign: implement conversion tracking, with the free capability provided by Google or via another method at your disposal.

Once you have that in place, look at how your AdWords campaign performs. If the conversion rate is a lofty 5% for one keyword and 3% for another keyword, chances are that your organic search conversion rates for those two keywords will vary by a similar amount. Be aware, though, that although paid search results get significantly less traffic than organic search results, paid click-throughs do tend to convert at a somewhat higher rate (1.25 to 1.5 times,

according to Enquisite). Using the preceding example, this suggests that we will get a little less than one conversion per day as a result of being in the #1 position.

This data is great to have in hand. However, it does not mean you should use this methodology instead of other keyword tools. It takes time to implement, and time to collect the data. The keyword tools will get you started in real time. Nonetheless, using AdWords, Yahoo! Search Marketing, or MSN adCenter can provide you with some great data.

Using Landing Page Optimization

Landing page optimization (sometimes also called *conversion optimization*) is the practice of actively testing multiple variations of a web page (or website) to see which one performs the best. Typically, this is done as part of an effort to improve the conversion performance of the site.

The simplest form of this type of test is called an A/B test. A/B tests involve creating two different versions of a page, and then randomly picking which version to show to a new visitor to the site (old visitors get the version they saw the last time they visited). You then measure the behavior of the visitors to the two different versions to see which group of visitors completes conversions on the site. You have to be careful to wait until you have a statistically significant amount of data to draw a conclusion. Once you have this data you can analyze it and decide on more tests, or simply pick the winner and move on.

Multivariate testing is a bit more complex, because it involves more than two variations in the test. In addition, you can mix and match multiple variations. For example, you may want to try two different logos, two different calls to action, three different page titles, two different color schemes, and so on. In multivariate testing, any combination of your elements could be what is shown to a particular visitor. Obviously, more data (visits and actions) is required to draw a conclusion than in a simple A/B test.

Landing page optimization can help in determining the value of a keyword because one of the elements you can be testing is the impact on conversion of variations of a keyphrase in the page title, the page header, and in other strategic places on the page. One variation of the test would use one keyphrase and the other variation would use a different one.

You can then see which keyword provides the best results. This data can provide you with an interesting measure of keyword value—its ability to help you convert your visitors. However, landing page optimization is not something you can use to perform SEO tests (i.e., to see which version of a page ranks higher), as SEO tests can take weeks or even months to see results.

Leveraging the Long Tail of Keyword Demand

As we discussed at the beginning of this chapter, the long tail of search is where 70% of search queries occur. Only 30% of those precious queries happen in the more obvious terms that people use, the so-called “head terms.” Another way to underscore this is that in May 2007,

Google Vice President Udi Manber indicated that 20% to 25% of all search queries that Google receives on a given day are queries that Google is seeing for the first time. You can think of this as the “ultra-long tail.”

The long tail of search queries in a given industry is typically not visible via any of the major keyword research services or search engine ad databases (Google AdWords, Yahoo! Search Marketing, and MSN adCenter). In these instances, there is a research method to find those terms that can carry value, but it requires a good amount of research and analysis.

With this in mind, let’s outline a few methods for finding long tail terms.

Extracting Terms from Relevant Web Pages

One source for long tail terms is web pages that do well for searches that are relevant to your target market. Here is a basic process for finding those pages and extracting that information from them:

1. Extract the top 10 to 50 most common search phrases at the head of the distribution graph from your existing keyword research in the industry.
2. Search Google, Yahoo!, and Bing for each term.
3. For each page in the top 10 to 30 results, extract the unique usable text on the page.
4. Remove stop words and filter by phrase size.
5. Remove instances of terms/phrases already in your keyword research database.
6. Sort through the most common remnants first, and comb as far down as you feel is valuable.

Through this process, you are basically text-mining relevant documents on the subject of your industry/service/product for terms that, although lower in search volume, have a reasonable degree of relation. When using this process, it is imperative to have human eyes reviewing the extracted data to make sure it passes the “common sense” test. You may even find additional terms at the head of the keyword distribution graph.

You can expand on this method in the following ways:

- Text-mine Technorati or Delicious for relevant results.
- Use documents purely from specific types of results—local, academic—to focus your keyword mining efforts.
- Mine forum threads on your subject matter. You could even use *inurl:forum* in the searches to grab conversational keywords.

This methodology is highly effective. The return on this research has a direct relationship to the amount of effort you expend (and how deep you dig).

Mining Keyword Research Tools

Although looking into keyword research tools for long tail data has significant limitations, there are still ways to do it. For example, if you own a chain of pizza restaurants in 50 cities across the country and you want to discover long tail terms, you can. Let's look at the tail end of Wordtracker's output for a combined search on *Orlando Pizza*, *San Diego Pizza*, and *San Jose Pizza* (see [Figure 5-35](#)).

Line 39, *pizza san diego delivery*, is an example of a valid long tail term. If some people search for *pizza san diego delivery*, it becomes quite likely that others search for *pizza orlando delivery*. It does not show in this data, because the volume of queries available to the keyword research tool is limited. All we are doing with these combined searches is giving the search tools more data to work with.

The takeaway remains valid: apply these logical long tail extensions across all of your cities, even though the keyword tool shows it for only one, and you're likely to attract search queries for those keywords.

Identifying Long Tail Patterns

You can also take another stab at determining long tail information. As a hypothetical example using *digital camera*, here are 40 searches for two different brands and models of digital cameras that have been pulled (for this demonstration) from the KeywordDiscovery database that received only one search:

- *consumer comments on nikon 5.1 mp coolpix l3 digital camera*
- *new nikon coolpix p3 8 1 mp digital camera memory*
- *nikon 3 2 mp coolpix digital camera*
- *nikon 51 mp coolpix s1 digital camera and cradle*
- *nikon 6 mp coolpix digital camera*
- *nikon 7 1 mp coolpix 7900 digital camera*
- *nikon 81 mp coolpix 8800 digital camera*
- *nikon coolpix 4800 4 mp digital camera*
- *nikon coolpix 5200 51 mp digital camera*
- *nikon coolpix 5400 51 mp digital camera*
- *nikon coolpix 6.0 mp digital camera*
- *nikon coolpix 8700 8mp 8x zoom digital camera 8 mp*
- *nikon coolpix l2 6.0 mp digital camera*
- *nikon coolpix l3 6 mp digital camera usa warranty*
- *nikon coolpix p2 51 mp digital camera*

- *best buy sony cybershot dsc t7 51 mp digital camera*
- *brand new sony cybershot dsc h1 51 mp digital camera*
- *camera digital sony cybershot 51 mp*
- *sony - cybershot 10.1 mp digital camera*
- *sony - cybershot 6.0 mp digital camera*
- *sony 5 mp cybershot dsc t9 digital camera*
- *sony 72 mp cybershot dsc p200 digital camera information*
- *sony 72 mp cybershot dsc w7 digital camera*
- *sony 72 mp digital still camera cybershot rebate*
- *sony cybershot 10.1 mp digital camera*
- *sony cybershot 7 2mp digital camera 7 2 mp*
- *sony cybershot 72mp dsc w7 digital camera 72 mp*
- *sony cybershot 81 mp digital camera*
- *sony cybershot digital camera 5.1 mp*
- *sony cybershot digital camera 6 mp*
- *sony cybershot dsc 1 81 mp digital camera review*
- *sony cybershot dsc h1 51 mp digital camera*
- *sony cybershot dsc w30 6 mp digital camera*
- *sony cybershot dscs40 41 mp digital camera 3x opt zoom*
- *sony dsc p73 cybershot digital camera 41 mp p 73*
- *sony dsc p8 cybershot 32 mp digital camera*
- *sony dsc s60 cybershot digital camera 4 1 mp*
- *sony dsc s85 cybershot 41 mp digital still camera*
- *sony dsc t1 cybershot digital camera 5 0 mp*
- *sony dsc t1 cybershot digital camera 50 mp t 1*

37	<input type="checkbox"/> sammy's woodfire pizza san diego
38	<input type="checkbox"/> sammy's woodire pizza san diego
39	<input type="checkbox"/> pizza san diego delivery
40	<input type="checkbox"/> anthony's pizza orlando fl
41	<input type="checkbox"/> broadway pizza san diego
42	<input type="checkbox"/> dominos pizza 9145 chesapeake dr san diego
43	<input type="checkbox"/> broadway pizza orlando fl
44	<input type="checkbox"/> anthony's pizza orlando
45	<input type="checkbox"/> anthony's coal fired pizza coupons orlando
46	<input type="checkbox"/> anthony's coal fired pizza coupons orlando
47	<input type="checkbox"/> dominos pizza phone number local orlando fl curryford
48	<input type="checkbox"/> gaetano's pizza san diego
49	<input type="checkbox"/> italian village pizza in orlando
50	<input type="checkbox"/> la bellas pizza in san diego
51	<input type="checkbox"/> incredible pizza orlando
52	<input type="checkbox"/> house of pizza orlando
53	<input type="checkbox"/> giovannis pizza in san diego

FIGURE 5-35. Extracting long tail data from Wordtracker

Our goal is to determine whether there are any universal patterns that searchers tend to use when searching. Within this subset of searches, a number of patterns stand out:

- Approximately 48% begin with the brand name and end with *digital camera*.
- Approximately 35% are ordered brand, model name, model number, megapixel, *digital camera*.
- Approximately 22.5% are ordered brand, megapixel, model name, *digital camera*.
- A whopping 60% follow the overall pattern of brand, model name, *digital camera*.

You might also notice that, at least in this example, qualifiers such as *new*, a specific store name, and a reference to consumer comments tend to precede the search phrases, whereas features and product-related qualifiers such as *memory*, *3x opt zoom*, *warranty*, *cradle*, *information*, and even a repeat of the megapixels or model number tend to be appended to the search phrases.

N O T E

Remember, this is purely a limited, hypothetical example and certainly is not meant to be statistically accurate. The goal here was to reveal different search term patterns to aid in determining the best groupings of long tail keywords to target.

Editorial Content Strategies for Long Tail Targeting

One of the most difficult aspects of capturing traffic from the long tail of search is creating relevant, targeted content. As we discussed in “[Determining Searcher Intent and Delivering Relevant, Fresh Content](#)” on page 41 in [Chapter 2](#), search engines rely on lexical analysis to determine what a web page is about. As a result, your chances of showing up for a long tail phrase are greatly increased if you have that long tail phrase, or at least all the words that make up the long tail phrase, on your page. Let’s look at why this may be challenging by checking out what phrases Wordtracker returns when we enter *canon digital camera* (see [Figure 5-36](#)).

□ canon digital camera	525	520
□ canon digital camera reviews	49	47
□ canon digital slr camera	26	25
□ canon powershot sd780 is digital camera	20	19
□ canon digital camera software	18	17
□ canon g10 digital camera	17	16
□ canon digital camera repair	16	15
□ canon digital camera accessories	15	14
□ canon powershot s1100is 8mp digital camera	13	12
□ canon powershot a590 is digital camera	13	12
□ canon digital camera solution disk	13	12
□ canon powershot a590 digital camera	10	9
□ canon powershot sd800 is digital elph camera silver	9	8
□ sony dsc-hx1 canon sx1 is digital camera review	9	8

FIGURE 5-36. Sample long tail data

Already, with the eleventh keyphrase returned (*canon digital camera solution disk*), you can see the challenge. If you are trying to sell Canon digital cameras you are probably not going to work that keyphrase into your page copy.

The best approach is to use the long tail research techniques we discussed in this chapter and identify the major patterns, or the major words that appear across different long tail scenarios, and then work those words into your copy. Don’t force it and make pages that appear foolish to a user.

Make sure the writers remain focused on producing quality content. From a long tail perspective, more text is better because it creates more possible long tail matches, but there are limits to that too. Don’t put a 1,000-word article on your site unless it makes sense to your users for you to do so.

User-Generated Content Strategies for Long Tail Targeting

User-generated content (UGC) can be a great way to obtain lots of content that will help attract long tail traffic. Popular ways of doing that include forums, reviews, blog comments, and a way to upload videos or images, among others. As users submit content, they do the hard work of writing the text you need to capitalize on the long tail.

There are some downsides to UGC, though. Generally speaking, you need to moderate it to make sure people are not contributing objectionable material you don't want on your site. Even if you get community members to participate, you will still need to manage them.

In addition, you need to have a strategy for getting the process started. In the case of a forum, you need to develop a critical mass of users to establish a real community. If you don't establish this critical mass, a high percentage of the posts you receive will be one form of spam or another. To make UGC work, you need one or more of the following:

- Significant existing daily site traffic. How much depends on how vertically oriented your community is intended to be. Narrowly focused topics can get going with a smaller number of users.
- A way to generate a lot of buzz to generate site traffic.
- Compelling supporting content.

If you can succeed at this, you'll give life to a machine that produces long tail content on an ongoing basis with comparatively low effort.

Trending, Seasonality, and Seasonal Fluctuations in Keyword Demand

One of the subtleties of keyword research, and of any fully developed SEO strategy, is that the use of keywords varies significantly over time. For instance, major holidays inevitably lead to bursts of keyword volume related to those holidays. Examples could be searches such as *Halloween costumes*, *gift ideas for Christmas*, or *Valentine's candy*.

If you want to write holiday-related content, it will be important to have your site visible in the SERPs for those search queries prior to that holiday's buying season so that you'll get optimum traffic for those terms. And since the search engines take considerable time in ranking your pages, advance preparation is required. To investigate this further, let's examine the Google Trends data for a period of 12 months for the search term *halloween costume ideas* (see [Figure 5-37](#)).

As you can see, searches begin gaining traction toward the end of August and into autumn; thus, if you are doing SEO for Halloween-related terms, you would want to have the related content and links in place by the beginning of the summer so that search engines can find and index your content, and therefore you're more visible to searchers when they start doing

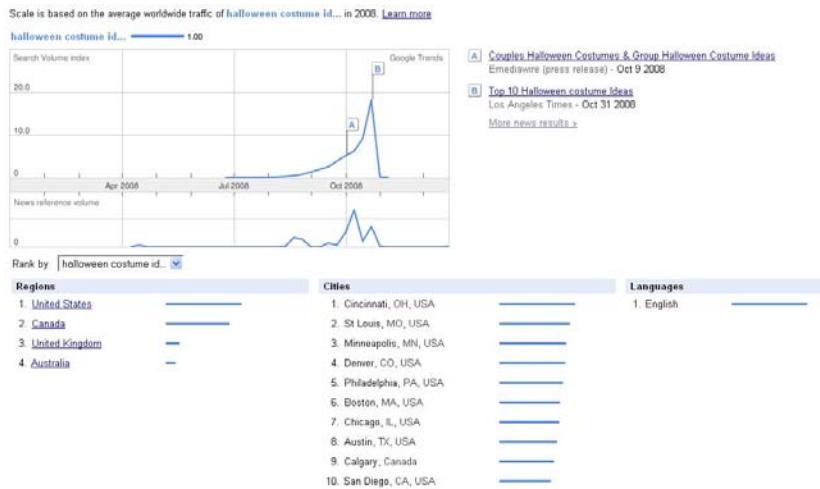


FIGURE 5-37. Google Trends highly seasonal data example

research. A long-term SEO approach would take this into consideration as part of the overall strategy for the site. Searches start consistently increasing toward the end of September. You can keep trying more examples. With Valentine's Day, the searches start in mid-December.

A similar pattern emerges for Christmas-related searches. Figure 5-38 shows an example for *christmas gift ideas*. Searches start consistently increasing toward the end of September. You can keep trying more examples. With Valentine's Day, the searches start in mid-December.

In each case, searches started increasing about two to three months before the holiday, so it is important to acknowledge that and start crafting your content and targeting those keywords in ample time for them to be indexed before the searches start gaining traction.

KeywordDiscovery (<http://www.keyworddiscovery.com>) also graphs search trends, so if you have an account, you can analyze these graphs to craft a holiday campaign, as shown in Figure 5-39.

You can see when people begin to search for Halloween costumes and when the activity drops off. Looking a bit farther down you can see the search engines receiving the search queries and the countries they are coming from, as shown in Figure 5-40.

Don't take your cue from when the stores start stocking Halloween candy—do the research and find out what last year's trends were so that you're prepared this year. If you prepare early enough, you'll be ready, while your competitors are scrambling with last-minute link-building campaigns three weeks before the holiday.

Also, don't remove your Halloween (or other seasonal) page as soon as the time frame has passed. Once you have fought hard to get rankings for your seasonal trophy term, you want to make sure you get the benefit for that hard work next year. Too many sites delete or archive these seasonal pages after the season is over, and then they have to start over again next year.

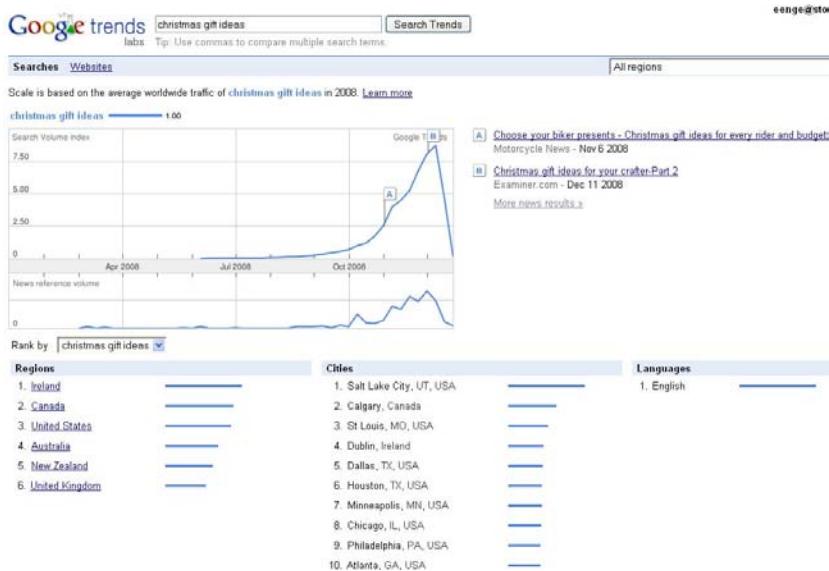


FIGURE 5-38. Google Trends; another seasonal example

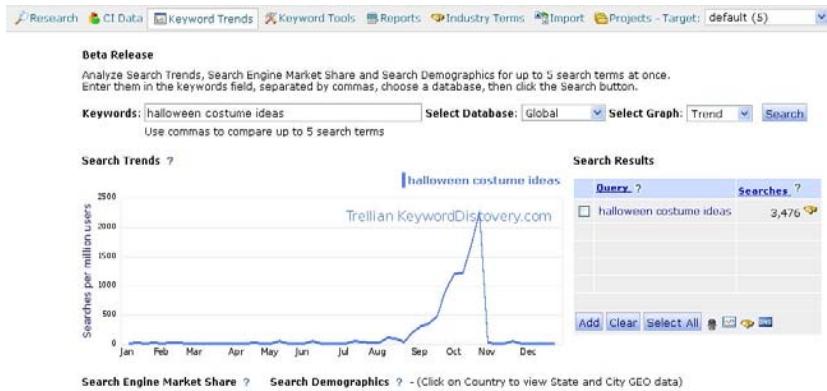


FIGURE 5-39. KeywordDiscovery seasonal data example

A better strategy is to leave the page in place until a new version is created, reuse the same URL, and archive the old content to a different URL. Leaving the page in place will give you a jumpstart when it is time to begin ramping up next year.



FIGURE 5-40. KeywordDiscovery keyword data by search engine

Conclusion

Keyword research is a complex and time-consuming topic, but the rewards are high. Once you learn where the keyword search volume is, you can begin to think about how that affects the information architecture and the navigation structure of your site. These are two critical elements that we will explore in detail in [Chapter 6](#).

CHAPTER SIX

Developing an SEO-Friendly Website

IN THIS CHAPTER, WE WILL EXAMINE THE MAJOR ELEMENTS OF HOW TO ASSESS THE search engine friendliness of your site. Having your site accessible to search engines enables search engines to discover what content is on your website. Then, when the search engines receive a search query, they can consider whether your content is a match for it.

As we discussed in the introduction to [Chapter 2](#), search engine crawlers are basically software programs. This provides them with certain strengths and weaknesses. Publishers must adapt their websites to make the job of these software programs easier—in essence, leverage their strengths and make their weaknesses irrelevant. If you can do this, you will have taken a major step forward toward success with SEO.

Developing an SEO-friendly site architecture requires a significant amount of communication and planning due to the large number of factors that influence the ways a search engine sees your site, and the large number of ways in which a website can be put together, as there are hundreds (if not thousands) of tools that web developers can use to build a website, many of which were not initially designed with SEO in mind.

Making Your Site Accessible to Search Engines

The first step is to ensure that your site can be found and crawled by search engines. This is not as simple as it sounds, as there are many popular web design and implementation constructs that the crawlers may not understand.

Indexable Content

To rank well in the search engines, your site's content—that is, the material available to visitors of your site—must be in HTML text form. Images, Flash files, Java applets, and other nontext content is, for the most part, virtually invisible to search engine spiders despite advances in crawling technology.

Although the easiest way to ensure that the words and phrases you display to your visitors are visible to search engines is to place the content in the HTML text on the page, more advanced methods are available for those who demand greater formatting or visual display styles. For example, images in GIF, JPEG, or PNG format can be assigned alt attributes in HTML, providing search engines with a text description of the visual content. Likewise, images can be shown to visitors as replacements for text by using CSS styles, via a technique called *CSS image replacement*.

Spiderable Link Structures

As we outlined in [Chapter 2](#), search engines use links on web pages to help them discover other web pages and websites. For this reason, website developers should invest the time to build a link structure that spiders can crawl easily. Many sites make the critical mistake of hiding or obfuscating their navigation in ways that make spiderability difficult, thus impacting their ability to get pages listed in the search engines' indexes. Consider the illustration in [Figure 6-1](#) that shows how this problem can occur.

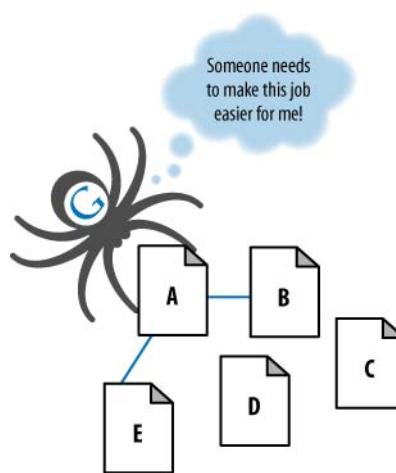


FIGURE 6-1. Providing search engines with crawlable link structures

In [Figure 6-1](#), Google's spider has reached Page A and sees links to pages B and E. However, even though pages C and D might be important pages on the site, the spider has no way to reach them (or even to know they exist) because no direct, crawlable links point to those pages.

As far as Google is concerned, they might as well not exist—great content, good keyword targeting, and smart marketing won’t make any difference at all if the spiders can’t reach those pages in the first place.

To refresh your memory on the discussion in [Chapter 2](#), here are some common reasons why pages may not be reachable:

Links in submission-required forms

Search spiders will not attempt to “submit” forms, and thus, any content or links that are accessible only via a form are invisible to the engines. This even applies to simple forms such as user logins, search boxes, or some types of pull-down lists.

Links in nonparsable JavaScript

If you use JavaScript for links, you may find that search engines either do not crawl or give very little weight to the links embedded within them.

Links in Flash, Java, or other plug-ins

Links embedded inside Java and plug-ins are invisible to the engines. In theory, the search engines are making progress in detecting links within Flash, but don’t rely too heavily on this.

Links in PowerPoint and PDF files

PowerPoint and PDF files are no different from Flash, Java, and plug-ins. Search engines sometimes report links seen in PowerPoint files or PDFs, but how much they count for is not easily known.

Links pointing to pages blocked by the meta Robots tag, rel="NoFollow", or robots.txt

The *robots.txt* file provides a very simple means for preventing web spiders from crawling pages on your site. Use of the *NoFollow* attribute on a link, or placement of the *meta Robots* tag on the page containing the link, is an instruction to the search engine to not pass link juice via the link (a concept we will discuss further in [“Content Delivery and Search Spider Control” on page 238](#)).

Links on pages with many hundreds or thousands of links

Google has a suggested guideline of 100 links per page before it may stop spidering additional links from that page. This “limit” is somewhat flexible, and particularly important pages may have upward of 150 or even 200 links followed. In general, however, it is wise to limit the number of links on any given page to 100 or risk losing the ability to have additional pages crawled.

Links in frames or iframes

Technically, links in both frames and iframes can be crawled, but both present structural issues for the engines in terms of organization and following. Unless you’re an advanced user with a good technical understanding of how search engines index and follow links in frames, it is best to stay away from them as a place to offer links for crawling purposes. We will discuss frames and iframes in more detail in [“Creating an Optimal Information Architecture” on page 187](#).

XML Sitemaps

Google, Yahoo!, and Microsoft all support a protocol known as XML Sitemaps. Google first announced it in 2005, and then Yahoo! and Microsoft agreed to support the protocol in 2006. Using the Sitemaps protocol you can supply the search engines with a list of all the URLs you would like them to crawl and index.

Adding a URL to a Sitemap file does not guarantee that a URL will be crawled or indexed. However, it can result in pages that are not otherwise discovered or indexed by the search engine getting crawled and indexed. In addition, Sitemaps appear to help pages that have been relegated to Google's supplemental index make their way into the main index.

This program is a complement to, not a replacement for, the search engines' normal, link-based crawl. The benefits of Sitemaps include the following:

- For the pages the search engines already know about through their regular spidering, they use the metadata you supply, such as the last date the content was modified (*lastmod date*) and the frequency at which the page is changed (*changefreq*), to improve how they crawl your site.
- For the pages they don't know about, they use the additional URLs you supply to increase their crawl coverage.
- For URLs that may have duplicates, the engines can use the XML Sitemaps data to help choose a canonical version.
- Verification/registration of XML Sitemaps may indicate positive trust/authority signals.
- The crawling/inclusion benefits of Sitemaps may have second-order positive effects, such as improved rankings or greater internal link popularity.

The Google engineer who in online forums goes by GoogleGuy (a.k.a. Matt Cutts, the head of Google's webspam team) has explained Google Sitemaps in the following way:

Imagine if you have pages A, B, and C on your site. We find pages A and B through our normal web crawl of your links. Then you build a Sitemap and list the pages B and C. Now there's a chance (but not a promise) that we'll crawl page C. We won't drop page A just because you didn't list it in your Sitemap. And just because you listed a page that we didn't know about doesn't guarantee that we'll crawl it. But if for some reason we didn't see any links to C, or maybe we knew about page C but the URL was rejected for having too many parameters or some other reason, now there's a chance that we'll crawl that page C.

Sitemaps use a simple XML format that you can learn about at <http://www.sitemaps.org>. XML Sitemaps are a useful and in some cases essential tool for your website. In particular, if you have reason to believe that the site is not fully indexed, an XML Sitemap can help you increase the number of indexed pages. As sites grow in size, the value of XML Sitemap files tends to increase dramatically, as additional traffic flows to the newly included URLs.

Layout of an XML Sitemap

The first step in the process of creating an XML Sitemap is to create an *.xml* Sitemap file in a suitable format. Since creating an XML Sitemap requires a certain level of technical know-how, it would be wise to involve your development team in the XML Sitemap generator process from the beginning. [Figure 6-2](#) shows an example of some code from a Sitemap.

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.google.com/schemas/sitemap/0.84">
  <url>
    <loc>http://www.example.com/</loc>
    <lastmod>2009-09-09</lastmod>
    <changefreq>monthly</changefreq>
    <priority>0.8</priority>
  </url>
</urlset>
```

FIGURE 6-2. Sample XML Sitemap from Google.com

To create your XML Sitemap, you can use the following:

An XML Sitemap generator

This is a simple script that you can configure to automatically create Sitemaps, and sometimes submit them as well. Sitemap generators can create these Sitemaps from a URL list, access logs, or a directory path hosting static files corresponding to URLs. Here are some examples of XML Sitemap generators:

- [SourceForge.net's google-sitemap_gen](#)
- [ROR Sitemap Generator](#)
- [XML-Sitemaps.com Sitemap Generator](#)
- [Sitemaps Pal](#)
- [XML Echo](#)

Simple text

You can provide Google with a simple text file that contains one URL per line. However, Google recommends that once you have a text Sitemap file for your site, you use the Sitemap Generator to create a Sitemap from this text file using the Sitemaps protocol.

Syndication feed

Google accepts Really Simple Syndication (RSS) 2.0 and Atom 1.0 feeds. Note that the feed may provide information on recent URLs only.

What to include in a Sitemap file

When you create a Sitemap file you need to take care to include only the canonical version of each URL. In other words, in situations where your site has multiple URLs that refer to one

piece of content, search engines may assume that the URL specified in a Sitemap file is the preferred form of the URL for the content. You can use the Sitemap file as one way to suggest to the search engines which is the preferred version of a given page.

In addition, be careful about what not to include. For example, do not include multiple URLs that point to identical content; leave out pages that are simply pagination pages, or alternate sort orders for the same content, and/or any low-value pages on your site. Plus, make sure that none of the URLs listed in the Sitemap file include any tracking parameters.

Where to upload your Sitemap file

When your Sitemap file is complete, upload the file to your site in the highest-level directory you want search engines to crawl (generally, the root directory). If you list URLs in your Sitemap that are at a higher level than your Sitemap location, the search engines will be unable to include those URLs as part of the Sitemap submission.

Managing and updating XML Sitemaps

Once your XML Sitemap has been accepted and your site has been crawled, monitor the results and update your Sitemap if there are issues. With Google, you can return to <http://www.google.com/webmasters/sitemaps/siteoverview> to view the statistics and diagnostics related to your Google Sitemaps. Just click the site you want to monitor. You'll also find some FAQs from Google on common issues such as slow crawling and low indexation.

Update your XML Sitemap with the big three search engines when you add URLs to your site. You'll also want to keep your Sitemap file up-to-date when you add a large volume of pages or a group of pages that are strategic.

There is no need to update the XML Sitemap when simply updating content on existing URLs. It is not strictly necessary to update when pages are deleted as the search engines will simply not be able to crawl them, but you want to update them before you have too many broken pages in your feed. With the next update after adding new pages, however, it is a best practice to also remove those deleted pages; make the current XML Sitemap as accurate as possible.

Updating your Sitemap with Bing. Simply update the *.xml* file in the same location as before.

Updating your Google Sitemap. You can resubmit your Google Sitemap using your Google Sitemaps account, or you can resubmit it using an HTTP request:

From Google Sitemaps

Sign into Google Webmaster Tools with your Google account. From the Sitemaps page, select the checkbox beside your Sitemap filename and click the Resubmit Selected button. The submitted date will update to reflect this latest submission.

From an HTTP request

If you do this, you don't need to use the Resubmit link in your Google Sitemaps account. The Submitted column will continue to show the last time you manually clicked the link,

but the Last Downloaded column will be updated to show the last time Google fetched your Sitemap. For detailed instructions on how to resubmit your Google Sitemap using an HTTP request, see <http://www.google.com/support/webmasters/bin/answer.py?answer=34609>.

Google and the other major search engines discover and index websites by crawling links. Google XML Sitemaps are a way to feed the URLs that you want crawled on your site to Google for more complete crawling and indexation, which results in improved long tail searchability. By creating and updating this *.xml* file, you are helping to ensure that Google recognizes your entire site, and this recognition will help people find your site. It also helps the search engines understand which version of your URLs (if you have more than one URL pointing to the same content) is the canonical version.

Creating an Optimal Information Architecture

Making your site friendly to search engine crawlers also requires that you put some thought into your site information architecture. A well-designed architecture can bring many benefits for both users and search engines.

The Importance of a Logical, Category-Based Flow

The search engines face myriad technical challenges in understanding your site. Crawlers are not able to perceive web pages in the way that humans do, and thus significant limitations for both accessibility and indexing exist. A logical and properly constructed website architecture can help overcome these issues and bring great benefits in search traffic and usability.

At the core of website organization are two critical principles: usability, or making a site easy to use; and information architecture, or crafting a logical, hierarchical structure for content.

One of the very early proponents of information architecture, Richard Saul Wurman, developed the following definition for *information architect*:

information architect. 1) the individual who organizes the patterns inherent in data, making the complex clear. 2) a person who creates the structure or map of information which allows others to find their personal paths to knowledge. 3) the emerging 21st century professional occupation addressing the needs of the age focused upon clarity, human understanding, and the science of the organization of information.

Usability and search friendliness

Search engines are trying to reproduce the human process of sorting relevant web pages by quality. If a real human were to do this job, usability and user experience would surely play a large role in determining the rankings. Given that search engines are machines and they don't have the ability to segregate by this metric quite so easily, they are forced to employ a variety

of alternative, secondary metrics to assist in the process. The most well known and well publicized among these is link measurement (see Figure 6-3), and a well-organized site is more likely to receive links.

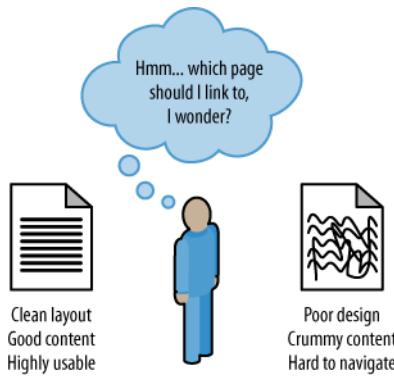


FIGURE 6-3. Making your site attractive to link to

Since Google launched in the late 1990s, search engines have strived to analyze every facet of the link structure on the Web and have extraordinary abilities to infer trust, quality, reliability, and authority via links. If you push back the curtain and examine why links between websites exist and how they come into place, you can see that a human being (or several humans, if the organization suffers from bureaucracy) is almost always responsible for the creation of links.

The engines hypothesize that high-quality links will point to high-quality content, and that great content and positive user experiences will be rewarded with more links than poor user experiences. In practice, the theory holds up well. Modern search engines have done a very good job of placing good-quality, usable sites in top positions for queries.

An analogy

Look at how a standard filing cabinet is organized. You have the individual cabinet, drawers in the cabinet, folders within the drawers, files within the folders, and documents within the files (see Figure 6-4).

There is only one copy of any individual document, and it is located in a particular spot. There is a very clear navigation path to get to it.

If you want to find the January 2008 invoice for a client (Amalgamated Glove & Spat), you would go to the cabinet, open the drawer marked Client Accounts, find the Amalgamated Glove & Spat folder, look for the Invoices file, and then flip through the documents until you come to the January 2008 invoice (again, there is only one copy of this; you won't find it anywhere else).

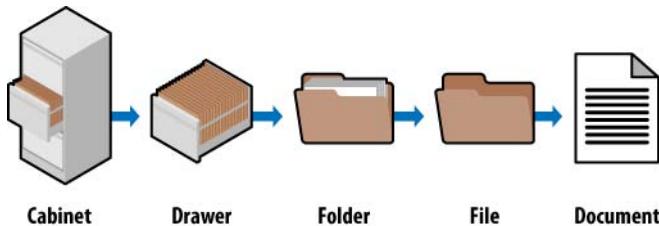


FIGURE 6-4. Similarities between filing cabinets and web pages

Figure 6-5 shows what it looks like when you apply this logic to the popular website, Craigslist.org.

Filing Category	Site Analogue
Cabinet	Fully Qualified Domain (e.g., seattle.craigslist.com)
Drawer	Primary Site Category (e.g., Housing)
Folder	Secondary Site Category (e.g., Apartments for Rent)
File	Tertiary Site Category (e.g., 2 Bedroom Apartments)
Document	Individual Page/Item (e.g., 2 Bedroom Capitol Hill Loft Apartment)

FIGURE 6-5. Filing cabinet analogy applied to Craigslist.org

If you're seeking an apartment on Capitol Hill in Seattle, you'd navigate to Seattle.Craigslist.org, choose Housing and then Apartments, narrow that down to two bedrooms, and pick the two-bedroom loft from the list of available postings. Craigslist's simple, logical information architecture has made it easy to reach the desired post in four clicks, without having to think too hard at any step about where to go. This principle applies perfectly to the process of SEO, where good information architecture dictates:

- As few clicks as possible to any given page
- One hundred or fewer links per page (so as not to overwhelm either crawlers or visitors)
- A logical, semantic flow of links from home page to categories to detail pages

Here is a brief look at how this basic filing cabinet approach can work for some more complex information architecture issues.

Subdomains. You should think of subdomains as completely separate filing cabinets within one big room. They may share similar architecture, but they shouldn't share the same content; and more importantly, if someone points you to one cabinet to find something, he is indicating that that cabinet is the authority, not the other cabinets in the room. Why is this important? It will help you remember that links (i.e., votes or references) to subdomains may not pass all, or any, of their authority to other subdomains within the room (e.g., “*.craigslist.com,” wherein “*” is a variable subdomain name).

Those cabinets, their contents, and their authority are isolated from each other and may not be considered to be in concert with each other. This is why, in most cases, it is best to have one large, well-organized filing cabinet instead of several that may prevent users and bots from finding what they want.

Redirects. If you have an organized administrative assistant, he probably uses 301 redirects inside his literal, metal filing cabinet. If he finds himself looking for something in the wrong place, he might place a sticky note in there reminding him of the correct location the next time he needs to look for that item. Anytime you looked for something in those cabinets, you could always find it because if you navigated improperly, you would inevitably find a note pointing you in the right direction. One copy. One. Only. Ever.

Redirect irrelevant, outdated, or misplaced content to the proper spot in your filing cabinet and both your users and the engines will know what qualities and keywords you think it should be associated with.

URLs. It would be tremendously difficult to find something in a filing cabinet if every time you went to look for it, it had a different name, or if that name resembled “jklhj25br3g452ikbr52k”. Static, keyword-targeted URLs are best for users and best for bots. They can always be found in the same place, and they give semantic clues as to the nature of the content.

These specifics aside, thinking of your site information architecture in terms of a filing cabinet is a good way to make sense of best practices. It'll help keep you focused on a simple, easily navigated, easily crawled, well-organized structure. It is also a great way to explain an often complicated set of concepts to clients and co-workers.

Since search engines rely on links to crawl the Web and organize its content, the architecture of your site is critical to optimization. Many websites grow organically and, like poorly planned filing systems, become complex, illogical structures that force people (and spiders) looking for something to struggle to find what they want.

Site Architecture Design Principles

In conducting website planning, remember that nearly every user will initially be confused about where to go, what to do, and how to find what he wants. An architecture that recognizes this difficulty and leverages familiar standards of usability with an intuitive link structure will have the best chance of making a visit to the site a positive experience. A well-organized site architecture helps solve these problems and provides semantic and usability benefits to both users and search engines.

In [Figure 6-6](#), a recipes website can use intelligent architecture to fulfill visitors' expectations about content and create a positive browsing experience. This structure not only helps humans navigate a site more easily, but also helps the search engines to see that your content fits into logical concept groups. You can use this approach to help you rank for applications of your product in addition to attributes of your product.

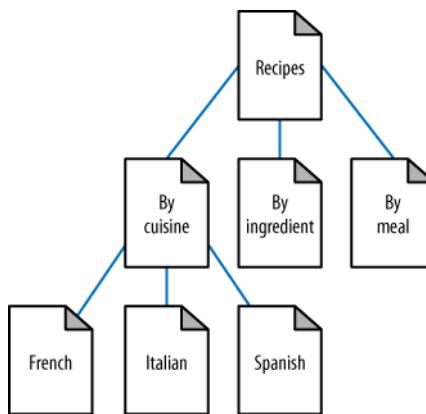


FIGURE 6-6. Structured site architecture

Although site architecture accounts for a small part of the algorithms, the engines do make use of relationships between subjects and give value to content that has been organized in a sensible fashion. For example, if in [Figure 6-6](#) you were to randomly jumble the subpages into incorrect categories, your rankings could suffer. Search engines, through their massive experience with crawling the Web, recognize patterns in subject architecture and reward sites that embrace an intuitive content flow.

Designing site architecture

Although site architecture—the creation structure and flow in a website's topical hierarchy—is typically the territory of information architects or is created without assistance from a company's internal content team, its impact on search engine rankings, particularly in the long

run, is substantial, thus making it wise to follow basic guidelines of search friendliness. The process itself should not be overly arduous, if you follow this simple protocol:

1. List all of the requisite content pages (blog posts, articles, product detail pages, etc.).
2. Create top-level navigation that can comfortably hold all of the unique types of detailed content for the site.
3. Reverse the traditional top-down process by starting with the detailed content and working your way up to an organizational structure capable of holding each page.
4. Once you understand the bottom, fill in the middle. Build out a structure for subnavigation to sensibly connect top-level pages with detailed content. In small sites, there may be no need for this level, whereas in larger sites, two or even three levels of subnavigation may be required.
5. Include secondary pages such as copyright, contact information, and other non-essentials.
6. Build a visual hierarchy that shows (to at least the last level of subnavigation) each page on the site.

[Figure 6-7](#) shows an example of a structured site architecture.

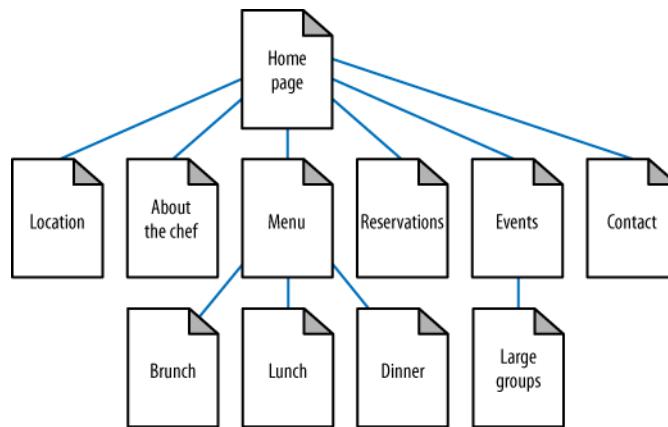


FIGURE 6-7. Second example of structured site architecture

Category structuring

As search engines crawl the Web, they collect an incredible amount of data (millions of gigabytes) on the structure of language, subject matter, and relationships between content. Though not technically an attempt at artificial intelligence, the engines have built a repository capable of making sophisticated determinations based on common patterns. As shown in [Figure 6-8](#), search engine spiders can learn semantic relationships as they crawl thousands of pages that cover a related topic (in this case, dogs).

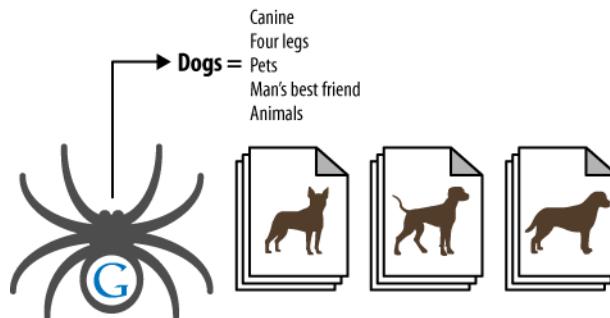


FIGURE 6-8. Spiders learning semantic relationships

Although content need not always be structured along the most predictable patterns, particularly when a different method of sorting can provide value or interest to a visitor, organizing subjects logically assists both humans (who will find your site easier to use) and engines (which will award you with greater rankings based on increased subject relevance).

Creating broad→narrow topical relevance. Naturally, this pattern of relevance-based scoring extends from single relationships between documents to the entire category structure of a website. Site creators can take advantage of this best by building hierarchies that flow from broad, encompassing subject matter down to more detailed, specific content. Obviously, in any categorization system, there's a natural level of subjectivity. Don't get too hung up on perfecting what the engines want here—instead, think first of your visitors and use these guidelines to ensure that your creativity doesn't overwhelm the project.

Taxonomy and ontology

In designing a website, you should also consider the taxonomy and ontology of the website. Taxonomy is essentially a two-dimensional hierarchical model of the architecture of the site. You can think of ontology as mapping the way the human mind thinks about a topic area. It can be much more complex than taxonomy, because a larger number of relationship types can be involved.

One effective technique for coming up with an ontology is called *card sorting*. This is a user-testing technique whereby users are asked to group items together so that you can organize your site as intuitively as possible. Card sorting can help identify not only the most logical paths through your site, but also ambiguous or cryptic terminology that should be reworded.

With card sorting, you write all the major concepts onto a set of cards that are large enough for participants to read, manipulate, and organize. Your test group assembles the cards in the order they believe provides the most logical flow, as well as into groups that seem to fit together.

By itself, building an ontology is not part of SEO, but when you do it properly it will impact your site architecture, and therefore it interacts with SEO. Coming up with the right site architecture should involve both disciplines.

Flat Versus Deep Architecture

One very strict rule for search friendliness is the creation of flat site architecture. Flat sites require a minimal number of clicks to access any given page, whereas deep sites create long paths of links required to access detailed content. For nearly every site with fewer than 10,000 pages, all content should be accessible through a maximum of three clicks from the home page and/or sitemap page. At 100 links per page, even sites with millions of pages can have every page accessible in five to six clicks if proper link and navigation structures are employed. If a site is not built to be flat, it can take too many clicks to reach the desired content, as shown in [Figure 6-9](#). In contrast, a flat site (see [Figure 6-10](#)) allows users and search engines to reach most content in just a few clicks.

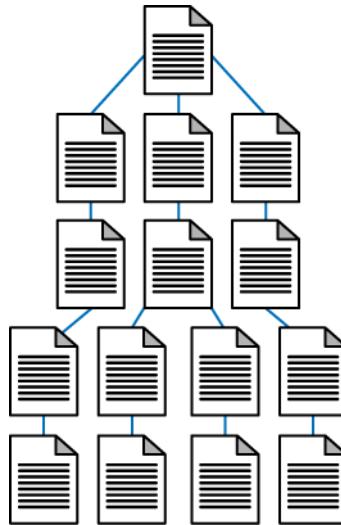


FIGURE 6-9. Deep site architecture

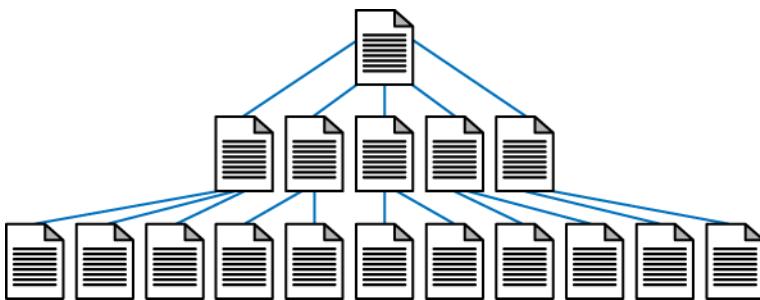


FIGURE 6-10. Flat site architecture

Flat sites aren't just easier for search engines to crawl; they are also simpler for users, as they limit the number of page visits the user requires to reach his destination. This reduces the abandonment rate and encourages repeat visits.

When creating flat sites, be aware that the engines are known to limit the number of links they crawl from a given page. Representatives from several of the major engines have said that more than 100 individual links from a single page might not be followed unless that page is of particular importance (i.e., many external sites link to it).

Following this guideline means limiting links on category pages to 100 or fewer, meaning that with the three-clicks-to-any-page rule, 10,000 pages is the maximum number possible (unless the home page is also used as a Sitemap-style link guide, which can increase the max, technically, to 1 million).

The 100-links-per-page issue relates directly to another rule for site architects: avoid pagination wherever possible. *Pagination*, the practice of creating a list of elements on pages separated solely by numbers (e.g., some e-commerce sites use pagination for product catalogs that have more products than they wish to show on a single page), is problematic for many reasons. First, pagination provides virtually no topical relevance. Second, content that moves into different pagination can create duplicate content issues. Last, pagination can create spider traps and hundreds or thousands of extraneous, low-quality pages that can be detrimental to search visibility. [Figure 6-11](#) shows how pagination structures do not benefit search engines.

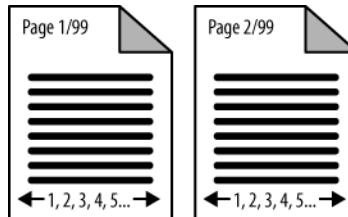


FIGURE 6-11. Pagination structures

So, make sure you implement flat structures and stay within sensible guidelines for the number of links per page, while retaining a contextually rich link structure. This is not always as easy as it sounds. Accomplishing this may require quite a bit of thought and planning to build a contextually rich structure on some sites. Consider a site with 10,000 different men's running shoes. Defining an optimal structure for that site could be a very large effort, but that effort will pay serious dividends in return.

Search-Friendly Site Navigation

Site navigation is something that web designers have been putting considerable thought and effort into since websites came into existence. Even before search engines were significant, navigation played an important role in helping users find what they wanted. It plays an important role in helping search engines understand your site as well.

Basics of search engine friendliness

The search engine spiders need to be able to read and interpret your website's code to properly spider and index the content on your web pages. Do not confuse this with the rules of organizations such as the W3C, which issues guidelines on HTML construction. Although following the W3C guidelines can be a good idea, the great majority of sites do not follow these guidelines, so search engines generally overlook violations of these rules as long as their spiders can parse the code.

Unfortunately, there are also a number of ways that navigation and content can be rendered on web pages that function for humans, but are invisible (or challenging) for search engine spiders.

For example, there are numerous ways to incorporate content and navigation on the pages of a website. For the most part, all of these are designed for humans. Basic HTML text and HTML links such as those shown in [Figure 6-12](#) work equally well for humans and search engine crawlers.

The text and the link that are indicated on the page shown in [Figure 6-12](#) (the Alchemist Media home page) are in simple HTML format.

Site elements that are problematic for spiders

However, many other types of content may appear on a web page and may work well for humans but not so well for search engines. Here are some of the most common ones.

Search and web forms

Many sites incorporate search functionality. These "site search" elements are specialized search engines that index and provide access to one site's content.



Search Engine Optimization (SEO) and Web Site Marketing Services: Organic Rankings & PPC Management

Alchemist Media has been featured for the past five years as one of BtoB Magazine's Top Search Engine Marketing companies in the U.S.

Links

[Company founder](#) Jessie Strichola is currently co-authoring "The Art of SEO" book with Rand Fishkin, Stephan Spencer, and Eric Enge, to be published by O'Reilly Media. She is one of the original nine founders of SEMPO (The Search Engine Marketing Professional Organization) and she spent two years on the Board of Directors for the organization.

Text

The organic search engine optimization techniques and web site marketing advice provided for you in this web site are based on over ten years of SEO and internet marketing experience. It is our hope that the lessons we have learned as search experts can help you understand the complicated task of search engine optimization within the ever-evolving world of search engine marketing.

Our company offers a variety of SEM and SEO services, including web site auditing for SEO friendliness/compatibility, SEO redesigns, and long term SEO Management, as well as PPC account management, web site usability analysis, and more. Our services range from hourly consulting to comprehensive annual engagements. In addition, we provide on-site trainings for companies with in-house SEO staff, and we are available for speaking engagements worldwide. We also provide discounted rates for non-profit organizations, and offer pro-bono arrangements in certain select circumstances.

FIGURE 6-12. Example page with simple text and text link

This is a popular method of helping users rapidly find their way around complex sites. For example, the [Pew Internet website](#) provides Site Search in the top-right corner; this is a great tool for users, but search engines will be stymied by it. Search engines operate by crawling the Web's link structure—they don't submit forms or attempt random queries into search fields, and thus, any URLs or content solely accessible via a "site search" function will remain invisible to Google, Yahoo!, and Bing.

Forms are a popular way to provide interactivity, and one of the simplest applications is the "contact us" form many websites have.

Unfortunately, crawlers will not fill out or submit forms such as these; thus, any content restricted to those who employ them is inaccessible to the engines. In the case of a "contact us" form, this is likely to have little impact, but other types of forms can lead to bigger problems.

Websites that have content behind logins will either need to provide text links to the content behind the login (which defeats the purpose of the login) or implement First Click Free (discussed in "[Content Delivery and Search Spider Control](#)" on page 238).

Java, images, audio, and video. Adobe Shockwave files, Java embeds, audio, and video (in any format) present content that is largely uncrawlable by the major engines. With some notable exceptions that we will discuss later, search engines can read text only when it is presented in HTML format. Embedding important keywords or entire paragraphs in an image or a Java console renders them invisible to the spiders. Likewise, words spoken in an audio file or video cannot be read by the search engines.

Alt attributes, originally created as metadata for markup and an accessibility tag for vision-impaired users, is a good way to present at least some text content to the engines when displaying images or embedded, nontext content. Note that the alt attribute is not a strong

signal, and using the `alt` attribute on an image link is no substitute for implementing a simple text link with targeted anchor text. A good alternative is to employ captions and text descriptions in the HTML content wherever possible.

In the past few years, a number of companies offering transcription services have cropped up, providing automated text creation for the words spoken in audio or video. Providing these transcripts on rich media pages makes your content accessible to the search engines and findable by keyword-searching visitors. You can also use software such as Dragon Naturally Speaking and dictate your “transcript” to your computer.

AJAX and JavaScript. JavaScript enables many dynamic functions inside a website, most of which interfere very minimally with the operations of a search engine spider. The exception comes when a page must use a JavaScript call to reach another page, or to pull content that the spiders can’t see in the HTML. Though these instances are relatively rare, it pays to be aware of how the robots spider and index—both content and links need to be accessible in the raw HTML of a page to avoid problems.

Asynchronous JavaScript and XML (AJAX) presents similar problems, most notably in the delivery of content that search engines cannot spider. Since AJAX uses database calls to retrieve data without refreshing a page or changing URLs, the content contained behind these technologies is frequently completely hidden from the search engines (see [Figure 6-13](#)).

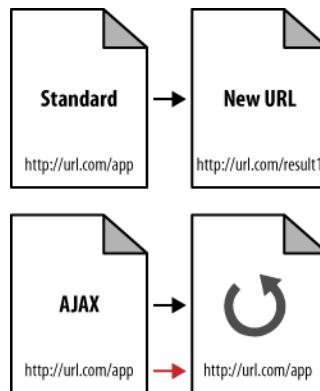


FIGURE 6-13. The problem with AJAX

When AJAX is used you may want to consider implementing an alternative spidering system for search engines to follow. AJAX applications are so user-friendly and appealing that for many publishers foregoing them is simply impractical. Building out a directory of links and pages that the engines can follow is a far better solution.

When you build these secondary structures of links and pages, make sure to provide users with access to them as well. Inside the AJAX application itself, give your visitors the option to

“directly link to this page” and connect that URL with the URL you provide to search spiders through your link structures. AJAX apps not only suffer from unspiderable content, but often don’t receive accurate links from users since the URL doesn’t change.

Newer versions of AJAX use a # delimiter, which acts as a query string into the AJAX application. This does allow you to link directly to different pages within the application. However, the #, which is used for HTML bookmarking, and everything past it, is ignored by search engines.

This is largely because web browsers use only what’s after the # to jump to the anchor within the page, and that’s done locally within the browser. In other words, the browser doesn’t send the full URL, so the parameter information (i.e., any text after the #) is not passed back to the server.

So, don’t use your ability to link to different pages within the AJAX application as a solution to the problem of exposing multiple pages within the application to search engines. All of the pages exposed in this way will be seen as residing on the same URL (everything preceding the #). Make sure you create discrete web pages that have unique URLs for the benefit of search engines.

Frames. Frames emerged in the mid-1990s as a popular way to make easy navigation systems. Unfortunately, both their usability (in 99% of cases) and their search friendliness (in 99.99% of cases) were exceptionally poor. Today, iframes and CSS can replace the need for frames, even when a site’s demands call for similar functionality.

For search engines, the biggest problem with frames and iframes is that they often hold the content from two or more URLs on a single page. For users, search engines, which direct searchers to only a single URL, may get confused by frames and direct visitors to single pages (orphan pages) inside a site intended to show multiple URLs at once.

Additionally, since search engines rely on links, and frame pages will often change content for users without changing the URL, external links often point to the wrong URL unintentionally. As a consequence, links to the page containing the frame or iframe may actually not point to the content the linker wanted to point to. [Figure 6-14](#) shows an example page that illustrates how multiple pages are combined into a single URL with frames, which results in link distribution and spidering issues.

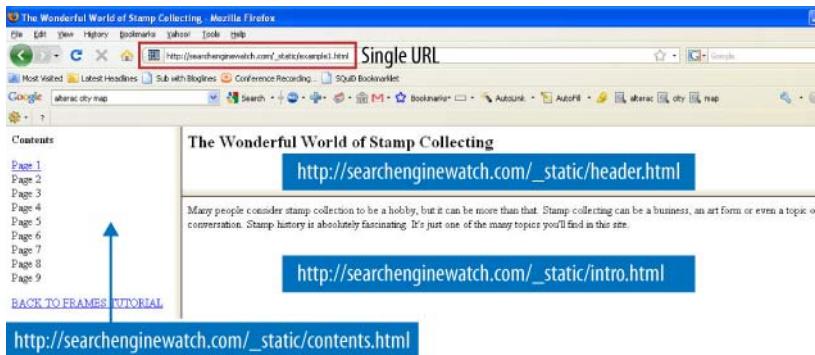


FIGURE 6-14. Sample page using frames

Search-engine-friendly navigation guidelines

Although search engine spiders have become more advanced over the years, the basic premise and goals remain the same: spiders find web pages by following links and record the content of the pages they find in the search engine's index (a giant repository of data about websites and pages).

In addition to avoiding the techniques we just discussed, there are some additional guidelines for developing search-engine-friendly navigation:

Implement a text-link-based navigational structure

If you choose to create navigation in Flash, JavaScript, or other technologies, make sure to offer alternative text links in HTML for spiders to ensure that automated robots (and visitors who may not have the required browser plug-ins) can reach your pages.

Beware of "spider traps"

Even intelligently coded search engine spiders can get lost in infinite loops of links that pass between pages on a site. Intelligent architecture that avoids looping 301 or 302 server codes (or other redirection protocols) should negate this issue, but sometimes online calendar links, infinite pagination that loops, or massive numbers of ways in which content is accessible or sorted can create tens of thousands of pages for search engine spiders when you intended to have only a few dozen true pages of content. You can read more about Google's viewpoint on this at <http://googlewebmastercentral.blogspot.com/2008/08/to-infinity-and-beyond-no.html>.

Watch out for session IDs and cookies

As we just discussed, if you limit the ability of a user to view pages or redirect based on a cookie setting or session ID, search engines may be unable to crawl your content. The bots do not have cookies enabled, nor can they deal with session IDs properly (each visit by the crawler gets a URL with a different session ID and the search engine sees these URLs

with session IDs as different URLs). Although restricting form submissions is fine (as search spiders can't submit forms anyway), limiting content access via cookies and session IDs is a bad idea. Does Google allow you to specify parameters in URLs? Yahoo! does. You can read more about it on seroundtable.com.

Server, hosting, and IP issues

Server issues rarely cause search engine ranking problems—but when they do, disastrous consequences can follow. The engines are acutely aware of common server problems, such as downtime or overloading, and will give you the benefit of the doubt (though this will mean your content cannot be spidered during periods of server dysfunction).

The IP address of your host can be of concern in some instances. IPs once belonging to sites that have spammed the search engines may carry with them negative associations that can hinder spidering and ranking. The engines aren't especially picky about shared hosting versus separate boxes, or about server platforms, but you should be cautious and find a host you trust.

Search engines have become paranoid about the use of certain domains, hosting problems, IP addresses, and blocks of IPs. Experience tells them that many of these have strong correlations with spam, and thus, removing them from the index can have great benefits for users. As a site owner *not* engaging in these practices, it pays to investigate your web host prior to getting into trouble.

You can read more about server and hosting issues in “[Identifying Current Server Statistics Software and Gaining Access](#)” on page 116 in Chapter 4.

Root Domains, Subdomains, and Microsites

Among the common questions in structuring a website (or restructuring one) are whether to host content on a new domain, when to use subfolders, and when to employ microsites.

As search engines scour the Web, they identify four kinds of web structures on which to place metrics:

Individual pages/URLs

These are the most basic elements of the Web—filenames, much like those that have been found on computers for decades, which indicate unique documents. Search engines assign query-independent scores—most famously, Google's PageRank—to URLs and judge them in their ranking algorithms. A typical URL might look something like <http://www.yourdomain.com/page.html>.

Subfolders

The folder structures that websites use can also inherit or be assigned metrics by search engines (though there's very little information to suggest that they are used one way or another). Luckily, they are an easy structure to understand. In the URL <http://www.yourdomain.com/blog/post17.html>, “/blog/” is the subfolder and “post17.html” is the name of the file in that subfolder. Engines may identify common features of documents

in a given subfolder and assign metrics to these (such as how frequently the content changes, how important these documents are in general, or how unique the content is that exists in these subfolders).

Subdomains/fully qualified domains (FQDs)/third-level domains

In the URL <http://blog.yourdomain.com/page.html>, three kinds of domain levels are present. The top-level domain (also called the *TLD* or *domain extension*) is “.com”, the second-level domain is “yourdomain”, and the third-level domain is “blog”. The third-level domain is sometimes referred to as a *subdomain*. Common web nomenclature does not typically apply the word *subdomain* when referring to *www*, although technically, this too is a subdomain. A fully qualified domain is the combination of the elements required to identify the location of the server where the content can be found (in this example, “*blog.yourdomain.com*”).

These structures can receive individual assignments of importance, trustworthiness, and value from the engines, independent of their second-level domains, particularly on hosted publishing platforms such as WordPress, Blogspot, Wetpaint, and so on.

Complete root domains/host domain/pay-level domains (PLDs)/second-level domains

The domain name you need to register and pay for, and the one you point DNS settings toward, is the second-level domain (though it is commonly improperly called the “top-level” domain). In the URL <http://www.yourdomain.com/page.html>, “yourdomain.com” is the second-level domain. Other naming conventions may refer to this as the “root” or “pay-level” domain.

Figure 6-15 shows some examples.

When to Use a Subfolder

If a subfolder will work it is the best choice 99.9% of the time. Keeping content on a single root domain and single subdomain (e.g., <http://www.yourdomain.com>) gives the maximum SEO benefits, as engines will maintain all of the positive metrics the site earns around links, authority, and trust, and will apply these to every page on the site.

Subfolders have all the flexibility of subdomains (the content *can*, if necessary, be hosted on a unique server or completely unique IP address through post-firewall load balancing) and none of the drawbacks. Subfolder content will contribute directly to how search engines (and users, for that matter) view the domain as a whole. Subfolders can be registered with the major search engine tools and geotargeted individually to specific countries and languages as well.

Although subdomains are a popular choice for hosting content, they are not recommended if SEO is a primary concern. Subdomains *may* inherit the ranking benefits and positive metrics of the root domain they are hosted underneath, but they do not always do so (and thus, content can underperform in these scenarios).



FIGURE 6-15. Breaking down some example URLs

When to Use a Subdomain

If your marketing team decides to promote a URL that is completely unique in content or purpose and would like to use a catchy subdomain to do it, using a subdomain can be practical. Sites such as Maps.google.com and Blog.searchenginewatch.com are examples of where the marketing considerations make a subdomain an acceptable choice. One good reason to use a subdomain is in a situation where using a subdomain can look more authoritative to users as a result of creating separation from the main domain.

Be wary of press and media attention to the domains, as unsavvy users often don't understand the concept of subdomains or that domains can be on the "World Wide Web" without a "www." It is much less expensive to use a subfolder and have slightly less marketing panache than it is to educate through branding and advertising.

Subdomains may also be a reasonable choice if keyword usage in the domain name is of critical importance. It appears that search engines do weight keyword usage in the URL, and have slightly higher benefits for exact matches in the subdomain (or third-level domain name) than subfolders.

When to Use a Separate Root Domain

If you have a single, primary site that has earned links, built content, and attracted brand attention and awareness, it is very rarely advisable to place any new content on a completely separate domain. There are rare occasions when this can make sense, and we'll walk through these, as well as explain how singular sites benefit from collecting all of their content in one root domain location.

Splitting similar or relevant content from your organization onto multiple domains can be likened to a store taking American Express Gold cards and rejecting American Express Corporate or American Express Blue—it is overly segmented and dangerous for the consumer mindset. If you can serve web content from a singular domain, that domain will earn branding among the minds of your visitors, references from them, links from other sites, and bookmarks from your regular customers. Switching to a new domain forces you to rebrand and to earn all of these positive metrics all over again.

Microsites

There is a lot of debate about microsites, and although we generally recommend that you do not saddle yourself with the hassle of dealing with multiple sites and their SEO risks and disadvantages, it is important to understand the arguments, if only a few, in favor of doing so.

Making the case for microsites

Optimized properly, a microsite may have dozens or even hundreds of pages. If your site is likely to gain more traction and interest with webmasters and bloggers by being at an arm's length from your main site, it may be worth considering—for example, if you have a very commercial main site, and you want to create some great content, perhaps as articles, podcasts, and RSS feeds, that do not fit on the main site.

When should you consider a microsite?

When you own a specific keyword search query domain

For example, if you own “[usedtoyotatrucks.com](#)”, you might do very well to pull in search traffic for the specific term *used toyota trucks* with a microsite.

When you plan to sell the domains

It is very hard to sell a folder or even a subdomain, so this strategy is understandable if you're planning to churn the domains in the second-hand market.

As discussed earlier, if you're a major brand building a “secret” or buzz-worthy microsite

In this case, it can be useful to use a separate domain (however, you really should 301 the pages of that domain back to your main site after the campaign is over so that the link juice continues to provide long-term benefit—just as the mindshare and branding do in the offline world).

You should never implement a microsite that acts as a doorway page to your main site, or that has substantially the same content as you have published on your main site. Consider this only if you are willing to invest in putting rich original content on the site, and if you are willing to invest the time to promote the site as an independent site.

Such a site may gain more links by being separated from the main commercial site. A microsite may have the added benefit of bypassing some of the legal and PR department hurdles and internal political battles. This could be a key consideration if you're at a monolithic or low-risk-tolerance organization.

However, a microsite on a brand-new domain may wallow in the Google sandbox for months (for more about the Google sandbox, see “[Determining Searcher Intent and Delivering Relevant, Fresh Content](#)” on page 41 in [Chapter 2](#)). So, what to do if you want to launch a microsite? Consider buying an aged, reputable domain “aftermarket”—one that has had a quality site on it for a while (parking pages don’t count!), and then change the domain registration information slowly so that the site’s PageRank doesn’t get reset to zero. Or start the clock running as soon as possible on your new domain by posting at least a few pages to the URL and then getting a few links to it—as far in advance of the official launch as possible.

Here are the reasons for not using a microsite:

Search algorithms favor large, authoritative domains

Take a piece of great content about a topic and toss it onto a small, mom and pop website—point some external links to it, optimize the page and the site for the target terms, and get it indexed. Now, take that exact same content and place it on Wikipedia or [CNN.com](#), or even SEOmoz—you’re virtually guaranteed that the content on the large, authoritative domain will outrank the content on the small niche site. The engines’ current algorithms favor sites that have built trust, authority, consistency, and history.

Multiple sites split the benefits of links

As suggested in [Figure 6-16](#), a single good link pointing to a page on a domain positively influences the entire domain and every page on it. Because of this phenomenon, it is much more valuable to have any link you can possibly get pointing to the same domain to help boost the rank and value of the pages on it. Having content or keyword-targeted pages on other domains that don’t benefit from the links you earn to your primary domain only creates more work.

100 links to Domain A ≠ 100 links to Domain B + 1 link to Domain A (from Domain B)

In [Figure 6-17](#), you can see how earning lots of links to Page G on a separate domain is far less valuable than earning those same links to a page on the primary domain. Due to this phenomenon, even if you interlink all of the microsites or multiple domains that you build, it still won’t be close to the value you can get from those links if they were to point directly to the primary domain.

**The rising tide (of a link)
lifts all boats (on the domain)**

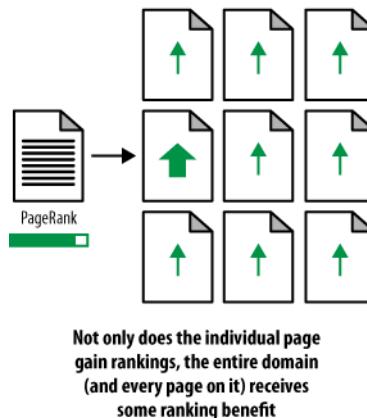


FIGURE 6-16. How links can benefit your whole site

A large, authoritative domain can host a huge variety of content

Niche websites frequently limit the variety of their discourse and content matter, whereas broader sites can target a wider range of foci. This is valuable not just for targeting the long tail of search and increasing potential branding and reach, but also for viral content, where a broader focus is much less limiting than that of a niche focus.

Time and energy are better spent on a single property

If you're going to pour your heart and soul into web development, design, usability, user experience, site architecture, SEO, public relations, branding, and so on, you want the biggest bang for your buck. Splitting your attention, time, and resources on multiple domains dilutes that value and doesn't let the natural order of building on your past successes on a single domain assist with that process. As shown in [Figure 6-16](#), every page on a site receives benefit from inbound links to a site. The page receiving the link gets the most benefit, but other pages also benefit.

When to Use a TLD Other Than .com

There are only a few rare situations in which you should consider using a TLD other than .com:

- When you own the .com and want to redirect to an .org, .tv, .biz, and so on, possibly for marketing/branding/geographic reasons. Do this only if you already own the .com and can redirect.

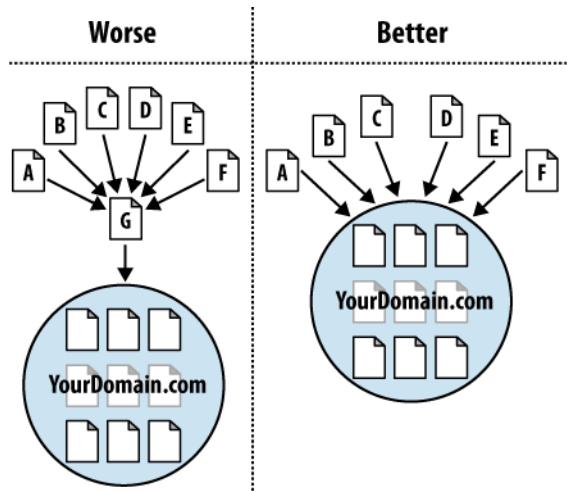


FIGURE 6-17. How direct links to your domain are better

- When you can use a .gov, .mil, or .edu domain (.jobs, though technically restricted to HR and hiring organizations, is available to anyone who hires and doesn't have any special search benefits).
- When you are serving only a single geographic region and are willing to permanently forego growth outside that region (e.g., .co.uk, .de, .it, etc.).
- When you are a nonprofit and want to distance your organization from the commercial world, .org may be for you.

Optimization of Domain Names/URLs

Two of the most basic parts of any website are the domain name and the URLs for the pages of the website. This section will explore guidelines for optimizing these important elements.

Optimizing Domains

When a new site is being conceived or designed one of the critical items to consider is the naming of the domain, whether it is for a new blog, a company launch, or even just a friend's website. Here are 12 tips that will be indispensable when helping you select a great domain name:

Brainstorm five top keywords

When you begin your domain name search, it helps to have five terms or phrases in mind that best describe the domain you're seeking. Once you have this list, you can start to pair them or add prefixes and suffixes to create good domain ideas. For example, if you're

launching a mortgage-related domain, you might start with words such as *mortgage*, *finance*, *home equity*, *interest rate*, and *house payment*, and then play around until you can find a good match.

Make the domain unique

Having your website confused with a popular site that someone else already owns is a recipe for disaster. Thus, never choose a domain that is simply the plural, hyphenated, or misspelled version of an already established domain. For example, Flickr desperately needs to buy [Flicker.com](#)—when kids in their 20s tell parents in their 40s and 50s to see photos on Flickr you can easily imagine that traffic going straight to the wrong domain.

Choose only dot-com available domains

If you're not concerned with type-in traffic, branding, or name recognition, you don't need to worry about this one. However, if you're at all serious about building a successful website over the long term, you should be worried about all of these elements, and although directing traffic to a .net or .org is fine, owning and 301'ing the .com is critical. With the exception of the very tech-savvy, most people who use the Web still make the automatic assumption that .com is all that's out there, or that these domains are more trustworthy. Don't make the mistake of locking out or losing traffic from these folks.

Make it easy to type

If a domain name requires considerable attention to type correctly due to spelling, length, or the use of unmemorable words or sounds, you've lost a good portion of your branding and marketing value. Usability folks even tout the value of having the words include easy-to-type letters (which we interpret as avoiding *q*, *z*, *x*, *c*, and *p*).

Make it easy to remember

Remember that word-of-mouth marketing relies on the ease with which the domain can be called to mind. You don't want to be the company with the terrific website that no one can ever remember to tell their friends about because they can't remember the domain name.

Keep the name as short as possible

Short names are easy to type and easy to remember (see the previous two rules). They also allow for more characters in the URL in the SERPs and a better fit on business cards and other offline media. Shorter URLs also get better click-through in the SERPs, according to a MarketingSherpa study located at <http://www.marketingsherpa.com/sample.cfm?ident=30181>.

Create and fulfill expectations

When someone hears about your domain name for the first time, he should be able to instantly and accurately guess at the type of content he might find there. That's why we love domain names such as [Hotmail.com](#), [CareerBuilder.com](#), [AutoTrader.com](#), and [WebMD.com](#). Domains such as [Monster.com](#), [Amazon.com](#), and [Zillow.com](#) required far more branding because of their nonintuitive names.

Avoid trademark infringement

This is a mistake that isn't made too often, but it can kill a great domain and a great company when it does. To be sure you're not infringing on anyone's registered trademark with your site's name, visit the [U.S. Patent and Trademark office site](#) and search before you buy.

Set yourself apart with a brand

Using a unique moniker is a great way to build additional value with your domain name. A "brand" is more than just a combination of words, which is why names such as [Mortgageforyourhome.com](#) and [Shoesandboots.com](#) aren't as compelling as branded names such as [Bankrate.com](#) and [Lendingtree.com](#).

Reject hyphens and numbers

Both hyphens and numbers make it hard to convey your domain name verbally and fall down on being easy to remember or type. Avoid spelled-out or Roman numerals in domains, as both can be confusing and mistaken for the other.

Don't follow the latest trends

Website names that rely on odd misspellings (as do many Web 2.0-style sites), multiple hyphens (such as the SEO-optimized domains of the early 2000s), or uninspiring short adjectives (such as "top...x," "best...x," "hot...x") aren't always the best choice. This isn't a hard and fast rule, but in the world of naming conventions in general, if everyone else is doing it, that doesn't mean it is a surefire strategy. Just look at all the people who named their businesses "AAA... x" over the past 50 years to be first in the phone book; how many Fortune 1000s are named "AAA company?"

Use an AJAX domain selection tool

Websites such as [Ajaxwhois](#) and [Domjax](#) make it exceptionally easy to determine the availability of a domain name. Just remember that you don't have to buy through these services. You can find an available name that you like, and then go to your registrar of choice.

Picking the Right URLs

Search engines place some weight on keywords in your URLs. Be careful, though. The search engines can interpret long URLs with numerous hyphens in them (e.g., [Buy-this-awesome-product-now.html](#)) as a spam signal. What follows are some guidelines for selecting optimal URLs for the pages of your site(s).

Describe your content

An obvious URL is a great URL. If a user can look at the Address bar (or a pasted link) and make an accurate guess about the content of the page before ever reaching it, you've done your job. These URLs get pasted, shared, emailed, written down, and yes, even recognized by the engines.

Keep it short

Brevity is a virtue. The shorter the URL, the easier to copy and paste, read over the phone, write on a business card, or use in a hundred other unorthodox fashions, all of which spell better usability and increased branding.

Static is the way

The search engines treat static URLs differently than dynamic ones. Users also are not fond of URLs in which the big players are ?, ?, and =. They are just harder to read and understand.

Descriptives are better than numbers

If you’re thinking of using “114/cat223/” you should go with “/brand/adidas/” instead. Even if the descriptive isn’t a keyword or particularly informative to an uninitiated user, it is far better to use words when possible. If nothing else, your team members will thank you for making it that much easier to identify problems in development and testing.

Keywords never hurt

If you know you’re going to be targeting a lot of competitive keyword phrases on your website for search traffic, you’ll want every advantage you can get. Keywords are certainly one element of that strategy, so take the list from marketing, map it to the proper pages, and get to work. For dynamically created pages through a CMS, create the option of including keywords in the URL.

Subdomains aren’t always the answer

First off, never use multiple subdomains (e.g., [Siteexplorer.search.yahoo.com](#)): they are unnecessarily complex and lengthy. Second, consider that subdomains have the potential to be treated separately from the primary domain when it comes to passing link and trust value. In most cases where just a few subdomains are used and there’s good interlinking, it won’t hurt, but be aware of the downsides. For more on this, and for a discussion of when to use subdomains, see “[Root Domains, Subdomains, and Microsites](#)” on page 201.

Fewer folders

A URL should contain no unnecessary folders (or words or characters, for that matter). They do not add to the user experience of the site and can in fact confuse users.

Hyphens separate best

When creating URLs with multiple words in the format of a phrase, hyphens are best to separate the terms (e.g., [/brands/dolce-and-gabbana/](#)), but you can also use plus signs (+).

Stick with conventions

If your site uses a single format throughout, don’t consider making one section unique. Stick to your URL guidelines once they are established so that your users (and future site developers) will have a clear idea of how content is organized into folders and pages. This can apply globally as well for sites that share platforms, brands, and so on.

Don’t be case-sensitive

Since URLs can accept both uppercase and lowercase characters, don’t ever, ever allow any uppercase letters in your structure. Unix/Linux-based web servers are case-sensitive,

so `http://www.domain.com/Products/widgets/` is technically a different URL from `http://www.domain.com/products/widgets/`. Note that this is not true in Microsoft IIS servers, but there are a lot of Apache web servers out there. In addition, this is confusing to users, and potentially to search engine spiders as well. If you have them now, 301 them to all-lowercase versions to help avoid confusion. If you have a lot of type-in traffic, you might even consider a 301 rule that sends any incorrect capitalization permutation to its rightful home.

Don't append extraneous data

There is no point in having a URL exist in which removing characters generates the same content. You can be virtually assured that people on the Web will figure it out, link to you in different fashions, confuse themselves, their readers, and the search engines (with duplicate content issues), and then complain about it.

Keyword Targeting

The search engines face a tough task; based on a few words in a query, sometimes only one, they must return a list of relevant results, order them by measures of importance, and hope that the searcher finds what he is seeking. As website creators and web content publishers, you can make this process massively simpler for the search engines and, in turn, benefit from the enormous traffic they send by employing the same terms users search for in prominent positions on your pages.

This practice has long been a critical part of search engine optimization, and although other metrics (such as links) have a great deal of value in the search rankings, keyword usage is still at the core of targeting search traffic.

The first step in the keyword targeting process is uncovering popular terms and phrases that searchers regularly use to find the content, products, or services your site offers. There's an art and science to this process, but it consistently begins with a list of keywords to target (see [Chapter 5](#) for more on this topic).

Once you have that list, you'll need to include these in your pages. In the early days of SEO, the process involved stuffing keywords repetitively into every HTML tag possible. Now, keyword relevance is much more aligned with the usability of a page from a human perspective.

Since links and other factors make up a significant portion of the search engines' algorithms, they no longer rank pages with 61 instances of "free credit report" above pages that contain only 60. In fact, *keyword stuffing*, as it is known in the SEO world, can actually get your pages devalued via search engine penalties. The engines don't like to be manipulated, and they recognize keyword stuffing as a disingenuous tactic. [Figure 6-18](#) shows an example of a page utilizing accurate keyword targeting.



FIGURE 6-18. *Title and headings tags—powerful for SEO*

Keyword usage includes creating titles, headlines, and content designed to appeal to searchers in the results (and entice clicks), as well as building relevance for search engines to improve your rankings.

Building a search-friendly site requires that the keywords searchers use to find content are prominently employed. Here are some of the more prominent places where a publisher can place those keywords.

Title Tags

For keyword placement, title tags are the most critical element for search engine relevance. The title tag is in the <head> section of an HTML document, and is the only piece of “meta” information about a page that influences relevancy and ranking. To give you an idea, when 72 well-known SEOs voted on what they believed were the most important ranking factors in Google’s algorithm, the majority of them said that a page’s title tag was the most important attribute (see SEOMoz’s search engine ranking factors study at <http://www.seomoz.org/article/search-ranking-factors> for more information).

The following eight rules represent best practices for title tag construction. Do keep in mind, however, that a title tag for any given page must directly correspond to that page’s content.

You may have five different keyword categories and a unique site page (or section) dedicated to each, so be sure to align a page's title tag content with its actual visible content as well.

Place your keywords at the beginning of the title tag

This provides the most search engine benefit, and thus, if you want to employ your brand name in the title tag, place it at the end. There is a tradeoff here between SEO benefit and branding benefit that you should think about and make explicitly. Major brands may want to place their brand at the start of the title tag as it may increase click-through rates. To decide which way to go you need to consider which need is greater for your business.

Limit length to 65 characters (including spaces)

Content in title tags after 65 characters is probably given less weight by the search engines. At a minimum, the title tag shown in the SERPs gets cut off at 65 characters. Watch this number carefully, though, as Google in particular is now supporting up to 70 characters in some cases.

Incorporate keyword phrases

This one may seem obvious, but it is critical to prominently include in your title tag whatever your keyword research shows as being the most valuable for capturing searches.

Target longer phrases if they are relevant

When choosing what keywords to include in a title tag, use as many as are completely relevant to the page at hand while remaining accurate and descriptive. Thus, it can be much more valuable to have a title tag such as "SkiDudes | Downhill Skiing Equipment & Accessories" rather than simply "SkiDudes | Skiing Equipment"—including those additional terms that are both relevant to the page and receive significant search traffic can bolster your page's value.

However, if you have separate landing pages for "skiing accessories" versus "skiing equipment," don't include one term in the other's title. You'll be cannibalizing your rankings by forcing the engines to choose which page on your site is more relevant for that phrase, and they might get it wrong. We will discuss the cannibalization issue in more detail shortly.

Use a divider

When splitting up the brand from the descriptive, options include | (a.k.a. the pipe), >, -, and ;, all of which work well. You can also combine these where appropriate—for example, "Major Brand Name: Product Category - Product". These characters do not bring an SEO benefit, but they can enhance the readability of your title.

Focus on click-through and conversion rates

The title tag is exceptionally similar to the title you might write for paid search ads, only it is harder to measure and improve because the stats aren't provided for you as easily. However, if you target a market that is relatively stable in search volume week to week, you can do some testing with your title tags and improve the click-through.

Watch your analytics and, if it makes sense, buy search ads on the page to test click-through and conversion rates of different ad text as well, even if it is for just a week or two. You can then look at those results and incorporate them into your titles, which can make a huge difference in the long run. A word of warning, though: don't focus entirely on click-through rates. Remember to continue measuring conversion rates.

Target searcher intent

When writing titles for web pages, keep in mind the search terms your audience employed to reach your site. If the intent is browsing or research-based, a more descriptive title tag is appropriate. If you're reasonably sure the intent is a purchase, download, or other action, make it clear in your title that this function can be performed at your site. Here is an example from <http://shopper.cnet.com>: "Buy Digital cameras - Best Digital camera prices - Shopper.com".

Be consistent

Once you've determined a good formula for your pages in a given section or area of your site, stick to that regimen. You'll find that as you become a trusted and successful "brand" in the SERPs, users will seek out your pages on a subject area and have expectations that you'll want to fulfill.

Meta Description Tags

Meta descriptions have three primary uses:

- To describe the content of the page accurately and succinctly
- To serve as a short text "advertisement" to click on your pages in the search results
- To display targeted keywords, not for ranking purposes, but to indicate the content to searchers

Great meta descriptions, just like great ads, can be tough to write, but for keyword-targeted pages, particularly in competitive search results, they are a critical part of driving traffic from the engines through to your pages. Their importance is much greater for search terms where the intent of the searcher is unclear or different searchers might have different motivations.

Here are seven good rules for meta descriptions:

Tell the truth

Always describe your content honestly. If it is not as "sexy" as you'd like, spice up your content; don't bait and switch on searchers, or they'll have a poor brand association.

Keep it succinct

Be wary of character limits—currently Google displays up to 160 characters, Yahoo! up to 165, and Bing up to 200+ (they'll go to three vertical lines in some cases). Stick with the smallest—Google—and keep those descriptions at 160 characters (including spaces) or less.

Author ad-worthy copy

Write with as much sizzle as you can while staying descriptive, as the perfect meta description is like the perfect ad: compelling and informative.

Test, refine, rinse, and repeat

Just like an ad, you can test meta description performance in the search results, but it takes careful attention. You'll need to buy the keyword through paid results (PPC ads) so that you know how many impressions critical keywords received over a given time frame. Then you can use analytics to see how many clicks you got on those keywords and calculate your click-through rate.

Analyze psychology

Unlike an ad, the motivation for a natural-search click is frequently very different from that of users clicking on paid results. Users clicking on PPC ads may be very directly focused on making a purchase, and people who click on a natural result may be more interested in research or learning about the company. Don't assume that successful PPC ad text will make for a good meta description (or the reverse).

Include relevant keywords

It is extremely important to have your keywords in the meta description tag—the boldface that the engines apply can make a big difference in visibility and click-through rate. In addition, if the user's search term is not in the meta description, chances are reduced that the meta description will be used as the description in the SERPs.

Don't employ descriptions universally

You shouldn't always write a meta description. Conventional logic may hold that it is usually wiser to write a good meta description yourself to maximize your chances of it being used in the SERPs, rather than let the engines build one out of your page content; however, this isn't always the case. If the page is targeting one to three heavily searched terms/phrases, go with a meta description that hits those users performing that search.

However, if you're targeting longer tail traffic with hundreds of articles or blog entries or even a huge product catalog, it can sometimes be wiser to let the engines themselves extract the relevant text. The reason is simple: when engines pull, they always display the keywords (and surrounding phrases) that the user searched for. If you try to force a meta description, you can detract from the relevance that the engines make naturally. In some cases, they'll overrule your meta description anyway, but since you can't consistently rely on this behavior, opting out of meta descriptions is OK (and for massive sites, it can save hundreds or thousands of man-hours).

Heading (H1, H2, H3) Tags

The H(x) tags in HTML (H1, H2, H3, etc.) are designed to indicate a headline hierarchy in a document. Thus, an H1 tag might be considered the headline of the page as a whole, whereas H2 tags would serve as subheadings, H3s as tertiary-level headlines, and so forth. The search



Forrester Releases Research on the Linkerati

Posted by [candish](#) on April 23rd, 2007 at 7:50 pm [Web Marketing](#)

Steve Rubel came back to blog just in time, [showcasing](#) this brilliant report from Forrester Research - [Social Technographics](#). The graphic he highlights is incredibly revealing on its own:

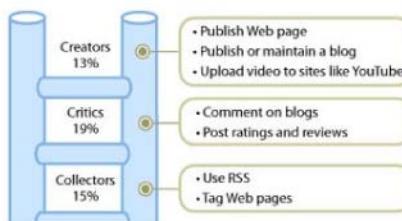


FIGURE 6-19. Headings styled to match the site

engines have shown a slight preference for keywords appearing in heading tags, notably the H1 tag (which is the most important of these to employ).

In some cases, you can use the title tag of a page, containing the important keywords, as the H1 tag. However, if you have a longer title tag, you may want to use a more focused, shorter heading tag using the most important keywords from the title tag. When a searcher clicks a result at the engines, reinforcing the search term he just typed in with the prominent headline helps to indicate that he has arrived on the right page with the same content he sought.

Many publishers assume that what makes the H1 a stronger signal is the size at which it is displayed. For the most part, the styling of your heading tags is not a factor in the SEO weight of the heading tag. You can style the tag however you want, as shown in [Figure 6-19](#), provided that you don't go to extremes (e.g., make it too small to read).

Document Text

The HTML text on a page was once the center of keyword optimization activities. Metrics such as keyword density and keyword saturation were used to measure the perfect level of keyword usage on a page. To the search engines, however, text in a document, particularly the frequency with which a particular term or phrase is used, has very little impact on how happy a searcher will be with that page.

In fact, quite often a page laden with repetitive keywords attempting to please the engines will provide a very poor user experience; thus, although some SEO professionals today do claim to use *term weight* (a mathematical equation grounded in the real science of information retrieval) or other, more “modern” keyword text usage methods, nearly all optimization can be done very simply.

The best way to ensure that you’ve achieved the greatest level of targeting in your text for a particular term or phrase is to use it in the title tag, in one or more of the section headings (within reason), and in the copy on the web page. Equally important is to use other related phrases within the body copy to reinforce the context and relevance of your main phrase for the page.

Although it is possible that implementing more instances of the key phrase on the page may result in some increase in ranking, this is increasingly unlikely to have an impact as you add more instances of the phrase. In addition, it can ruin the readability of some documents, which could hurt your ability to garner links to your site. Furthermore, testing has shown that document text keyword usage is such a small factor with the major engines that even one link of very low quality is enough to outweigh a page with perfect keyword optimization versus one that simply includes the targeted phrase naturally on the page (two to 10 times, depending on length).

This doesn’t mean keyword placement on pages is useless—you should always strive to include the keyword you’re targeting at least a few times depending on document length—but it does mean that aiming for “perfect” optimization on every page for every term is overkill and largely unnecessary.

Image Filenames and Alt Attributes

Incorporation of images on web pages can substantively enrich the user experience. However, the search engines cannot read the images directly. There are two elements that you can control to give the engines context for images:

The filename

Search engines look at the image filename to see whether it provides any clues to the content of the image. Don’t name your image *example.com/img4137a-b12.jpg*, as it tells the search engine nothing at all about the image, and you are passing up the opportunity to include keyword-rich text.

If it is a picture of Abe Lincoln, name the file *abe-lincoln.jpg* and/or have the SRC URL string contain it, as in example.com/abe-lincoln/portrait.jpg.

Image alt text

Image tags in HTML permit you to specify an attribute known as the `alt` attribute. This is a place where you can provide more information about what is in the image, and again where you can use your targeted keywords. Here is an example for the picture of Abe Lincoln:

```

```

Use the quotes if you have spaces in the text string of the `alt` content! Sites that have invalid `img` tags frequently lump a few words without quotes into the `img` tag, intended for the `alt` content—but with no quotes, all terms after the first word will be lost.

This usage of the image filename and of the `alt` attribute permits you to reinforce the major keyword themes of the page. This is particularly useful if you want to rank in image search. Make sure the filename and the `alt` text reflect the content of the picture, and do not artificially emphasize keywords unrelated to the image (even if they are related to the page). Although the `alt` attribute and the image filename are helpful, you should not use image links as a substitute for text links with rich anchor text that carries much more weight from an SEO perspective.

Presumably, your picture will relate very closely to the content of the page, and using the image filename and the `alt` text will help reinforce the page's overall theme.

Boldface Text

Some SEO professionals who engage in considerable on-page optimization testing have noticed that, all else being equal, a page that employs the targeted keyword(s) in `` or `` tags (HTML elements that boldface text visually) outrank their unbolted counterparts. Thus, although this is undoubtedly a very small factor in modern SEO, it may be worth leveraging, particularly for those looking to eke every last bit of optimization out of keyword usage.

Avoiding Keyword Cannibalization

As we discussed earlier, you should not use common keywords across multiple page titles. This advice applies to more than just the title tags.

One of the nastier problems that often crops up during the course of a website's information architecture, *keyword cannibalization* refers to a site's targeting of popular keyword search phrases on multiple pages, forcing the engines to pick which one is most relevant. In essence, a site employing cannibalization competes with itself for rankings and dilutes the ranking power of internal anchor text, external links, and keyword relevancy.

Avoiding cannibalization requires strict site architecture with attention to detail. Plot out your most important terms on a visual flowchart (or in a spreadsheet file, if you prefer), and pay careful attention to what search terms each page is targeting. Note that when pages feature two-, three-, or four-word phrases that contain the target search phrase of another page, linking back to that page within the content with the appropriate anchor text will avoid the cannibalization issue.

For example, if you had a page targeting “mortgages” and another page targeting “low-interest mortgages,” you would link back to the “mortgages” page from the “low-interest mortgages” page using the anchor text “mortgages” (see [Figure 6-20](#)). You can do this in the breadcrumb or in the body copy. The *New York Times* (<http://www.nytimes.com>) does the latter, where keywords in the body copy link to the related resource page on the site.

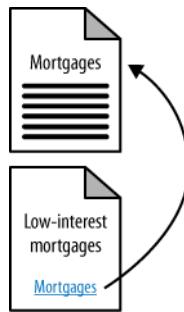


FIGURE 6-20. Adding lots of value with relevant cross-links

Keyword Targeting in CMSs and Automatically Generated Content

Large-scale publishing systems, or those that produce automatically generated content, present some unique challenges. If hundreds of pages are being created every day, it is not feasible to do independent keyword research on each and every page, making page optimization an interesting challenge.

In these scenarios, the focus turns to methods/recipes for generating unique titles, `<h1>` headings, and page content for each page. It is critical to educate the writers on ways to implement titles and headings that capture unique, key aspects of the articles’ content. More advanced teams can go further with this and train their writing staff on the use of keyword research tools to further optimize this process.

In the case of automatically generated material (such as that produced from algorithms that mine data from larger textual bodies), the key is to automate means for extracting a short (fewer than 70 characters) description of the article and making it unique from other titles generated elsewhere on the site and on the Web at large.

SEO Copywriting: Encouraging Effective Keyword Targeting by Content Creators

Very frequently, someone other than an SEO professional is responsible for content creation. Content creators often do not have an innate knowledge as to why keyword targeting is important—and therefore, training for effective keyword targeting is a critical activity. This is particularly important when dealing with large websites and large teams of writers.

Here are the main components of SEO copywriting that your writers must understand:

- Search engines look to match up a user's search queries with the keyword phrases on your web pages. If the search phrases do not appear on the page, chances are good that your page will never achieve significant ranking for that search phrase.
- The search phrases users may choose to use when looking for something are infinite in variety, but certain phrases will be used much more frequently than others.
- Using the more popular phrases you wish to target on a web page in the content for that page is essential to SEO success for that page.
- The title tag is the most important element on the page. Next up is the first header (H1), and then the main body of the content.
- Tools exist (as outlined in [Chapter 5](#)) that allow you to research and determine what the most interesting phrases are.

If you can get these five points across, you are well on your way to empowering your content creators to perform solid SEO. The next key element is training them on how to pick the right keywords to use.

This can involve teaching them how to use keyword research tools similar to the ones we discussed in [Chapter 5](#), or having the website's SEO person do the research and provide the terms to the writer.

The most important factor to reiterate to the content creator is that content quality and user experience still come first. Then, by intelligently making sure the right keyphrases are properly used throughout the content, they can help bring search engine traffic to your site. Reverse these priorities and you can end up with keyword stuffing or other spam issues.

Long Tail Keyword Targeting

As we outlined in [Chapter 5](#), the small-volume search terms, when tallied up, represent 70% of all search traffic, and the more obvious, high-volume terms represent only 30% of the overall search traffic.

For example, if you run a site targeting searches for *new york pizza* and *new york pizza delivery*, you might be surprised to find that hundreds of single searches each day for terms such as *pizza delivery on the corner of 57th & 7th*, or *Manhattan's tastiest Italian-style sausage pizza*, when taken

together, will actually provide considerably more traffic than the popular phrases you've researched. This concept is called the *long tail of search*.

Targeting the long tail is another aspect of SEO that combines art and science. In [Figure 6-21](#), you may not want to implement entire web pages for a history of pizza dough, pizza with white anchovies, or Croatian pizza.

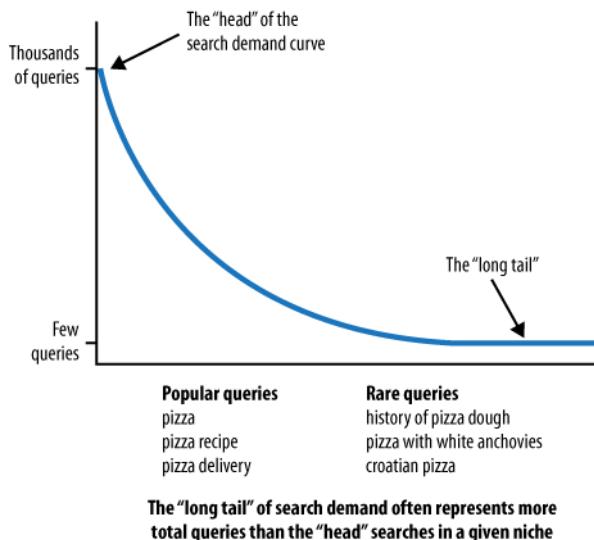


FIGURE 6-21. Example of the long tail search curve

Finding scalable ways to chase long tail keywords is a complex topic. Perhaps you have a page for ordering pizza in New York City, and you have a good title and H1 header on the page (e.g., “New York City Pizza: Order Here”), as well as a phone number and a form for ordering the pizza, and no other content. If that is all you have, that page is not competing effectively for rankings on long tail search terms. To fix this, you need to write additional content for the page. Ideally, this would be content that talks about the types of pizza that are popular in New York City, the ingredients used, and other things that might draw in long tail search traffic.

If you have a page for San Jose pizza, the picture gets even more complicated. You don't really want your content on the San Jose page to be the same as it is on the New York City page. This runs into potential duplicate content problems, as we will outline in [“Duplicate Content Issues” on page 226](#), or the keyword cannibalization issues we discussed earlier in this chapter.

To maximize your success, find a way to generate different content for those two pages, ideally tuned to the specific needs of the audience that arrives at those pages. Perhaps the pizza preferences of the San Jose crowd are different from those in New York City. Of course, the geographic information is inherently different between the two locations, so driving directions from key locations might be a good thing to include on the page.

If you have pizza parlors in 100 cities, this can get very complex indeed. The key here is to remain true to the diverse needs of your users, yet use your knowledge of the needs of search engines and searcher behavior to obtain that long tail traffic.

Content Optimization

Content optimization relates to how the presentation and architecture of the text, image, and multimedia content on a page can be optimized for search engines. Many of these recommendations are second-order effects. Having the right formatting or display won't boost your rankings directly, but through it, you're more likely to earn links, get clicks, and eventually benefit in search rankings. If you regularly practice the techniques in this section, you'll earn better consideration from the engines and from the human activities on the Web that influence their algorithms.

Content Structure

Because SEO has become such a holistic part of website improvement, it is no surprise that *content formatting*—the presentation, style, and layout choices you select for your content—is a part of the process. Choosing sans serif fonts such as Arial and Helvetica is a wise choice for the Web; Verdana in particular has received high praise from usability/readability experts, such as that which WebAIM offered in an article posted at <http://www.webaim.org/techniques/fonts/>.

Verdana is one of the most popular of the fonts designed for on-screen viewing. It has a simple, straightforward design, and the characters or glyphs are not easily confused. For example, the uppercase *I* and the lowercase *L* have unique shapes, unlike in Arial, in which the two glyphs may be easily confused (see [Figure 6-22](#)).



FIGURE 6-22. Arial versus Verdana font comparison

Another advantage of Verdana is the amount of spacing between letters. One consideration to take into account with Verdana is that it is a relatively large font. The words take up more space than words in Arial, even at the same point size (see [Figure 6-23](#)).

The larger size improves readability but also has the potential of disrupting carefully planned page layouts.

Verdana: Notice the length of this sentence.

Arial: Notice the length of this sentence.

FIGURE 6-23. How fonts impact space requirements

Font choice is accompanied in importance by sizing and contrast issues. Type that is smaller than 10 points is typically very challenging to read, and in all cases, relative font sizes are recommended so that users can employ browser options to increase/decrease if necessary. Contrast—the color difference between the background and text—is also critical; legibility usually drops for anything that isn’t black (or very dark) on a white background.

Content length and word count

Content length is another critical piece of the optimization puzzle that’s mistakenly placed in the “keyword density” or “unique content” bucket of SEO. In fact, content length can have a big role to play in terms of whether your material is easy to consume and easy to share. Lengthy pieces often don’t fare particularly well on the Web (with the exception, perhaps, of the one page sales letter), whereas short-form and easily digestible content often has more success. Sadly, splitting long pieces into multiple segments frequently backfires, as abandonment increases while link attraction decreases. The only benefit is page views per visit (which is why many sites which get their revenue from advertising employ this tactic).

Visual layout

Last but not least in content structure optimization is the display of the material. Beautiful, simplistic, easy-to-use, and consumable layouts instill trust and garner far more readership and links than poorly designed content wedged between ad blocks that threaten to overtake the page. For more on this topic, you might want to check out “The Golden Ratio in Web Design” from NetTuts (<http://nettuts.com/tutorials/other/the-golden-ratio-in-web-design/>), which has some great illustrations and advice on laying out web content on the page.

CSS and Semantic Markup

CSS is commonly mentioned as a best practice for general web design and development, but its principles provide some indirect SEO benefits as well. Google used to recommend keeping pages smaller than 101 KB, and it used to be a common belief that there were benefits to implementing pages that were small in size. Now, however, search engines deny that code size is a factor at all, unless it is really extreme. Still, keeping file size low means faster load times, lower abandonment rates, and a higher probability of being fully read and more frequently linked to.

CSS can also help with another hotly debated issue: code to text ratio. Some SEO professionals swear that making the code to text ratio smaller (so there’s less code and more text) can help

considerably on large websites with many thousands of pages. Your experience may vary, but since good CSS makes it easy, there's no reason not to make it part of your standard operating procedure for web development. Use tableless CSS stored in external files, keep JavaScript calls external, and separate the content layer from the presentation layer as shown on [CSS Zen Garden](#), a site that offers many user-contributed style sheets formatting the same HTML content.

Finally, CSS provides an easy means for “semantic” markup. For a primer, see Digital Web Magazine’s article, “Writing Semantic Markup” (http://www.digital-web.com/articles/writing_semantic_markup/). For SEO purposes, only a few primary tags apply, and the extent of microformats interpretation (using tags such as <author> or <address>) is less critical (the engines tend to sort out semantics largely on their own since so few web publishers participate in this coding fashion, but there is evidence that it helps with local search). Using CSS code to provide emphasis, to quote/reference, and to reduce the use of tables and other bloated HTML mechanisms for formatting, however, can make a positive difference.

Content Uniqueness and Depth

Few can debate the value the engines place on robust, unique, value-added content—Google in particular has had several rounds of kicking “low-quality-content” sites out of its indexes, and the other engines have followed suit.

The first critical designation to avoid is “thin content”—an insider phrase that (loosely) refers to content the engines do not feel contributes enough unique material to display a page competitively in the search results. The criteria have never been officially listed, but many examples/discussions from engineers and search engine representatives would place the following on the list:

- Thirty to 50 unique words, forming unique, parseable sentences that other sites/pages do not have.
- Unique HTML text content, different from other pages on the site in more than just the replacement of key verbs and nouns (yes, this means all those sites that build the same page and just change the city and state names thinking it is “unique” are mistaken).
- Unique titles and meta description elements. If you can’t write unique meta descriptions, just exclude them. Similarly, algorithms can trip up pages and boot them from the index simply for having near-duplicate meta tags.
- Unique video/audio/image content. The engines have started getting smarter about identifying and indexing pages for vertical search that wouldn’t normally meet the “uniqueness” criteria.

N O T E

By the way, you can often bypass these limitations if you have a good quantity of high-value external links pointing to the page in question (though this is very rarely scalable) or an extremely powerful, authoritative site (note how many one-sentence Wikipedia stub pages still rank).

The next criterion from the engines demands that websites “add value” to the content they publish, particularly if it comes from (wholly or partially) a secondary source.

A word of caution to affiliates

This word of caution most frequently applies to affiliate sites whose republishing of product descriptions, images, and so forth has come under search engine fire numerous times. In fact, it is best to anticipate manual evaluations here even if you’ve dodged the algorithmic sweep. The basic tenets are:

- Don’t simply republish something that’s found elsewhere on the Web unless your site adds substantive value to users, and don’t infringe on others’ copyrights.
- If you’re hosting affiliate content, expect to be judged more harshly than others, as affiliates in the SERPs are one of users’ top complaints about search engines.
- Small things such as a few comments, a clever sorting algorithm or automated tags, filtering, a line or two of text, simple mashups, or advertising do *not* constitute “substantive value.”

For some exemplary cases where websites fulfill these guidelines, check out the way sites such as [CNET](#), [Urbanspoon](#), and [Metacritic](#) take content/products/reviews from elsewhere, both aggregating *and* “adding value” for their users.

Last but not least, Google has provided a guideline to refrain from trying to place “search results in the search results.” For reference, look at the post from Google’s Matt Cutts, including the comments, at <http://www.mattcutts.com/blog/search-results-in-search-results/>. Google’s stated feeling is that search results generally don’t “add value” for users, though others have made the argument that this is merely an anticompetitive move.

Sites can benefit from having their “search results” transformed into “more valuable” listings and category/subcategory landing pages. Sites that have done this have had great success recovering rankings and gaining traffic from Google.

In essence, you want to avoid the potential for your site pages being perceived, both by an engine’s algorithm and by human engineers and quality raters, as search results. Refrain from:

- Pages labeled in the title or headline as “search results” or “results”

- Pages that appear to offer a query-based list of links to “relevant” pages on the site without other content (add a short paragraph of text, an image, and formatting that make the “results” look like detailed descriptions/links instead)
- Pages whose URLs appear to carry search queries (e.g., ?q=miami+restaurants or ?search=Miami+restaurants versus /miami-restaurants)
- Pages with text such as “Results 1 through 10”

Though it seems strange, these subtle, largely cosmetic changes can mean the difference between inclusion and removal. Err on the side of caution and dodge the appearance of search results.

Duplicate Content Issues

Duplicate content can result from many causes, including licensing of content to or from your site, site architecture flaws due to non-SEO-friendly CMSs, or plagiarism. Over the past five years, however, spammers in desperate need of content began the now much-reviled process of scraping content from legitimate sources, scrambling the words (through many complex processes), and repurposing the text to appear on their own pages in the hopes of attracting long tail searches and serving contextual ads (and various other nefarious purposes).

Thus, today we’re faced with a world of “duplicate content issues” and “duplicate content penalties.” Here are some definitions that are useful for this discussion:

Unique content

This is written by humans, is completely different from any other combination of letters, symbols, or words on the Web, and is clearly not manipulated through computer text-processing algorithms (such as Markov-chain-employed spam tools).

Snippets

These are small chunks of content such as quotes that are copied and reused; these are almost never problematic for search engines, especially when included in a larger document with plenty of unique content.

Shingles

Search engines look at relatively small phrase segments (e.g., five to six words) for the presence of the same segments on other pages on the Web. When there are too many shingles in common between two documents, the search engines may interpret them as duplicate content.

Duplicate content issues

This is typically used when referring to duplicate content that is not in danger of getting a website penalized, but rather is simply a copy of an existing page that forces the search engines to choose which version to display in the index (a.k.a. duplicate content filter).

Duplicate content filter

This is when the search engine removes substantially similar content from a search result to provide a better overall user experience.

Duplicate content penalty

Penalties are applied rarely and only in egregious situations. Engines may devalue or ban other web pages on the site, too, or even the entire website.

Consequences of Duplicate Content

Assuming your duplicate content is a result of innocuous oversights on your developer's part, the search engine will most likely simply filter out all but one of the pages that are duplicates because the search engine wants to display one version of a particular piece of content in a given SERP. In some cases, the search engine may filter out results prior to including them in the index, and in other cases the search engine may allow a page in the index and filter it out when it is assembling the SERPs in response to a specific query. In this latter case, a page may be filtered out in response to some queries and not others.

Searchers want diversity in the results, not the same results repeated again and again. Search engines therefore try to filter out duplicate copies of content, and this has several consequences:

- A search engine bot comes to a site with a crawl budget, which is counted in the number of pages it plans to crawl in each particular session. Each time it crawls a page that is a duplicate (which is simply going to be filtered out of search results) you have let the bot waste some of its crawl budget. That means fewer of your "good" pages will get crawled. This can result in fewer of your pages being included in the search engine index.
- Links to duplicate content pages represent a waste of link juice. Duplicated pages can gain PageRank, or link juice, and since it does not help them rank, that link juice is misspent.
- No search engine has offered a clear explanation for how its algorithm picks which version of a page it does show. In other words, if it discovers three copies of the same content, which two does it filter out? Which one does it still show? Does it vary based on the search query? The bottom line is that the search engine might not favor the version you wanted.

Although some SEO professionals may debate some of the preceding specifics, the general structure will meet with near-universal agreement. However, there are a couple of problems around the edge of this model.

For example, on your site you may have a bunch of product pages and also offer print versions of those pages. The search engine might pick just the printer-friendly page as the one to show in its results. This does happen at times, and it can happen even if the printer-friendly page has lower link juice and will rank less well than the main product page.

The fix for this is to apply the canonical URL tag to all versions of the page to indicate which version is the original.

Phase I: Google finds duplicate content

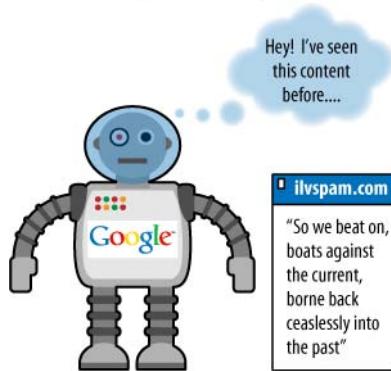


FIGURE 6-24. Google finding duplicate content

A second version of this can occur when you syndicate content to third parties. The problem is that the search engine may boot your copy of the article out of the results in favor of the version in use by the person republishing your article. The best fix for this, other than NoIndexing the copy of the article that your partner is using, is to have the partner implement a link back to the original source page on your site. Search engines nearly always interpret this correctly and emphasize your version of the content when you do that.

How Search Engines Identify Duplicate Content

Some examples will illustrate the process for Google as it finds duplicate content on the Web. In the examples shown in Figures 6-24 through 6-27, three assumptions have been made:

- The page with text is assumed to be a page containing duplicate content (not just a snippet, despite the illustration).
- Each page of duplicate content is presumed to be on a separate domain.
- The steps that follow have been simplified to make the process as easy and clear as possible. This is almost certainly not the exact way in which Google performs (but it conveys the effect).

There are a few facts about duplicate content that bear mentioning as they can trip up webmasters who are new to the duplicate content issue:

Location of the duplicate content

Is it duplicated content if it is all on my site? Yes, in fact, duplicate content can occur within a site or across different sites.

Phase II: Google checks comparable docs

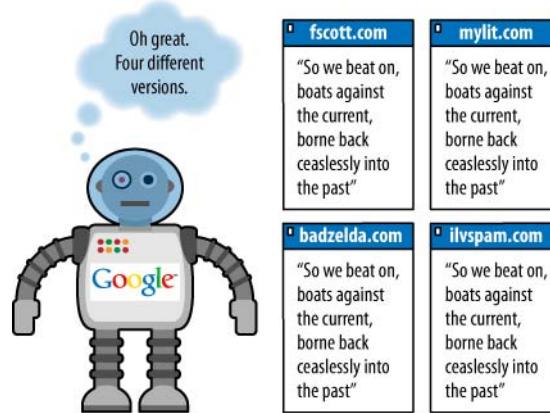


FIGURE 6-25. Google comparing the duplicate content to the other copies

Phase III: Duplicates get tossed out

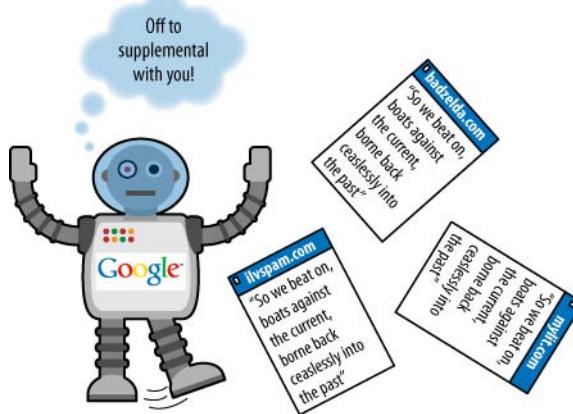


FIGURE 6-26. Duplicate copies getting tossed out

Percentage of duplicate content

What percentage of a page has to be duplicated before you run into duplicate content filtering? Unfortunately, the search engines would never reveal this information because it would compromise their ability to prevent the problem.

Phase IV: Google determines an original

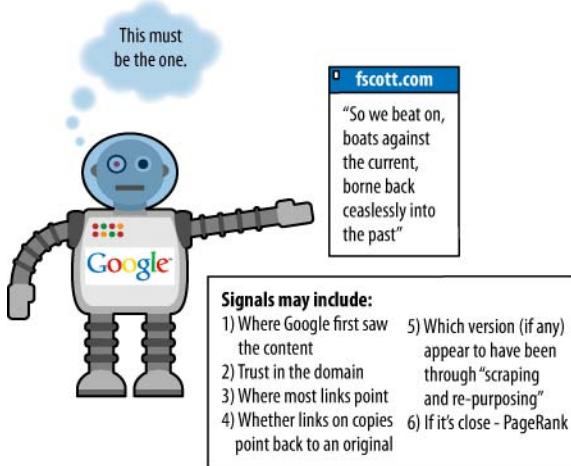


FIGURE 6-27. Google choosing one as the original

It is also a near certainty that the percentage at each engine fluctuates regularly and that more than one simple direct comparison goes into duplicate content detection. The bottom line is that pages do not need to be identical to be considered duplicates.

Ratio of code to text

What if your code is huge and there are very few unique HTML elements on the page? Will Google think the pages are all duplicates of one another? No. The search engines do not really care about your code; they are interested in the content on your page. Code size becomes a problem only when it becomes extreme.

Ratio of navigation elements to unique content

Every page on my site has a huge navigation bar, lots of header and footer items, but only a little bit of content; will Google think these pages are duplicates? No. Google (and Yahoo! and Bing) factor out the common page elements such as navigation before evaluating whether a page is a duplicate. They are very familiar with the layout of websites and recognize that permanent structures on all (or many) of a site's pages are quite normal. Instead, they'll pay attention to the "unique" portions of each page and often will largely ignore the rest.

Licensed content

What should I do if I want to avoid duplicate content problems, but I have licensed content from other web sources to show my visitors? Use `meta name = "robots" content="noindex, follow"`. Place this in your page's header and the search engines will know that the content isn't for them. This is a general best practice, because then humans can still visit the page, link to it, and the links on the page will still carry value.

Another alternative is to make sure you have exclusive ownership and publication rights for that content.

Identifying and Addressing Copyright Infringement

One of the best ways to monitor whether your site's copy is being duplicated elsewhere is to use [CopyScape.com](#), a site that enables you to instantly view pages on the Web that are using your content. Do not worry if the pages of these sites are in the supplemental index or rank far behind your own pages for any relevant queries—if any large, authoritative, content-rich domain tried to fight all the copies of its work on the Web, it would have at least two 40-hour-per-week jobs on its hands. Luckily, the search engines have placed trust in these types of sites to issue high-quality, relevant, worthy content, and therefore recognize them as the original issuer.

If, on the other hand, you have a relatively new site or a site with few inbound links, and the scrapers are consistently ranking ahead of you (or someone with a powerful site is stealing your work), you've got some recourse. One option is to file a DMCA infringement request with Google, with Yahoo!, and with Bing (you should also file this request with the site's hosting company).

The other option is to file a legal suit (or threaten such) against the website in question. If the site republishing your work has an owner in your country, this latter course of action is probably the wisest first step. You may want to try to start with a more informal communication asking them to remove the content before you send a letter from the attorneys, as the DMCA motions can take months to go into effect; but if they are nonresponsive, there is no reason to delay taking stronger action, either.

An actual penalty situation

The preceding examples show duplicate content filters and are not actual penalties, but, for all practical purposes, they have the same impact as a penalty: lower rankings for your pages. But there are scenarios where an actual penalty can occur.

For example, sites that aggregate content from across the Web can be at risk, particularly if little unique content is added from the site itself. In this type of scenario, you might see the site actually penalized.

The only fixes for this are to reduce the number of duplicate pages accessible to the search engine crawler, either by deleting them or `NoIndexing` the pages themselves, or to add a substantial amount of unique content.

One example of duplicate content that may get filtered out on a broad basis is a *thin affiliate* site. This nomenclature frequently describes a site promoting the sale of someone else's products (to earn a commission), yet provides little or no new information. Such a site may have received the descriptions from the manufacturer of the products and simply replicated

those descriptions along with an affiliate link (so that it can earn credit when a click/purchase is performed).

Search engineers have observed user data suggesting that, from a searcher's perspective, these sites add little value to their indexes. Thus, the search engines attempt to filter out this type of site, or even ban it from their index. Plenty of sites operate affiliate models but also provide rich new content, and these sites generally have no problem. It is when duplication of content and a lack of unique, value-adding material come together on a domain that the engines may take action.

How to Avoid Duplicate Content on Your Own Site

As we outlined, duplicate content can be created in many ways. Internal duplication of material requires specific tactics to achieve the best possible results from an SEO perspective. In many cases, the duplicate pages are pages that have no value to either users or search engines. If that is the case, try to eliminate the problem altogether by fixing the implementation so that all pages are referred to by only one URL. Also, 301-redirect the old URLs to the surviving URLs to help the search engines discover what you have done as rapidly as possible, and preserve any link juice the removed pages may have had.

If that process proves to be impossible, there are many options, as we will outline in "[Content Delivery and Search Spider Control](#)" on page 238. Here is a summary of the guidelines on the simplest solutions for dealing with a variety of scenarios:

- Use the canonical tag. This is the next best solution to eliminating the duplicate pages.
- Use *robots.txt* to block search engine spiders from crawling the duplicate versions of pages on your site.
- Use the Robots NoIndex meta tag to tell the search engine to not index the duplicate pages.
- NoFollow all the links to the duplicate pages to prevent any link juice from going to those pages. If you do this, it is still recommended that you NoIndex those pages as well.

You can sometimes use these tools in conjunction with one another. For example, you can NoFollow the links to a page and also NoIndex the page itself. This makes sense because you are preventing the page from getting link juice from your links, and if someone else links to your page from another site (which you can't control), you are still ensuring that the page does not get into the index.

However, if you use *robots.txt* to prevent a page from being crawled, be aware that using NoIndex or NoFollow on the page itself does not make sense, as the spider can't read the page, so it will never see the NoIndex or NoFollow tag. With these tools in mind, here are some specific duplicate content scenarios:

HTTPS pages

If you make use of *SSL* (encrypted communications between the browser and the web server often used for e-commerce purposes), you will have pages on your site that begin with https: instead of http:. The problem arises when the links on your https: pages link back to other pages on the site using relative instead of absolute links, so (for example) the link to your home page becomes *https://www.yourdomain.com* instead of *http://www.yourdomain.com*.

If you have this type of issue on your site, you may want to use the canonical URL tag, which we describe in “[Content Delivery and Search Spider Control](#) on page 238, or 301 redirects to resolve problems with these types of pages. An alternative solution is to change the links to absolute links (*http://www.yourdomain.com/content.html* instead of “/content.html”), which also makes life more difficult for content thieves that scrape your site.

CMSs that create duplicate content

Sometimes sites have many versions of identical pages because of limitations in the CMS where it addresses the same content with more than one URL. These are often unnecessary duplications with no end-user value, and the best practice is to figure out how to eliminate the duplicate pages and 301 the eliminated pages to the surviving pages. Failing that, fall back on the other options listed at the beginning of this section.

Print pages or multiple sort orders

Many sites offer print pages to provide the user with the same content in a more printer-friendly format. Or some e-commerce sites offer their products in multiple sort orders (such as size, color, brand, and price). These pages do have end-user value, but they do not have value to the search engine and will appear to be duplicate content. For that reason, use one of the options listed previously in this subsection.

Duplicate content in blogs and multiple archiving systems (pagination, etc.)

Blogs present some interesting duplicate content challenges. Blog posts can appear on many different pages, such as the home page of the blog, the Permalink page for the post, date archive pages, and category pages. Each instance of the post represents duplicates of the other instances. Once again, the solutions listed earlier in this subsection are the ones to use in addressing this problem.

User-generated duplicate content (repostings, etc.)

Many sites implement structures for obtaining user-generated content, such as a blog, forum, or job board. This can be a great way to develop large quantities of content at a very low cost. The challenge is that users may choose to submit the same content on your site and in several other sites at the same time, resulting in duplicate content among those sites. It is hard to control this, but there are two things you can do to reduce the problem:

- Have clear policies that notify users that the content they submit to your site must be unique and cannot be, or cannot have been, posted to other sites. This is difficult to enforce, of course, but it will still help some to communicate your expectations.

- Implement your forum in a different and unique way that demands different content. Instead of having only the standard fields for entering data, include fields that are likely to be unique over what other sites do, but that will still be interesting and valuable for site visitors to see.

Controlling Content with Cookies and Session IDs

Sometimes you want to more carefully dictate what a search engine robot sees when it visits your site. In general, search engine representatives will refer to the practice of showing different content to users than crawlers as *cloaking*, which violates the engines' Terms of Service (TOS) and is considered spam.

However, there are legitimate uses for this concept that are not deceptive to the search engines or malicious in intent. This section will explore methods for doing this with cookies and sessions IDs.

What's a Cookie?

A *cookie* is a small text file that websites can leave on a visitor's hard disk, helping them to track that person over time. Cookies are the reason Amazon.com remembers your username between visits and the reason you don't necessarily need to log in to your Hotmail account every time you open your browser. Cookie data typically contains a short set of information regarding when you last accessed a site, an ID number, and, potentially, information about your visit (see [Figure 6-28](#)).

Website developers can create options to remember visitors using cookies for tracking purposes or to display different information to users based on their actions or preferences. Common uses include remembering a username, maintaining a shopping cart, and keeping track of previously viewed content. For example, if you've signed up for an account with SEOmoz, it will provide you with options on your My Account page about how you want to view the blog and will remember that the next time you visit.

What Are Session IDs?

Session IDs are virtually identical to cookies in functionality, with one big difference. Upon closing your browser (or restarting), session ID information is no longer stored on your hard drive (usually); see [Figure 6-29](#). The website you were interacting with may remember your data or actions, but it cannot retrieve session IDs from your machine that don't persist (and session IDs by default expire when the browser shuts down). In essence, session IDs are more like temporary cookies (although, as you'll see shortly, there are options to control this).

Although technically speaking, session IDs are just a form of cookie without an expiration date, it is possible to set session IDs with expiration dates similar to cookies (going out decades). In

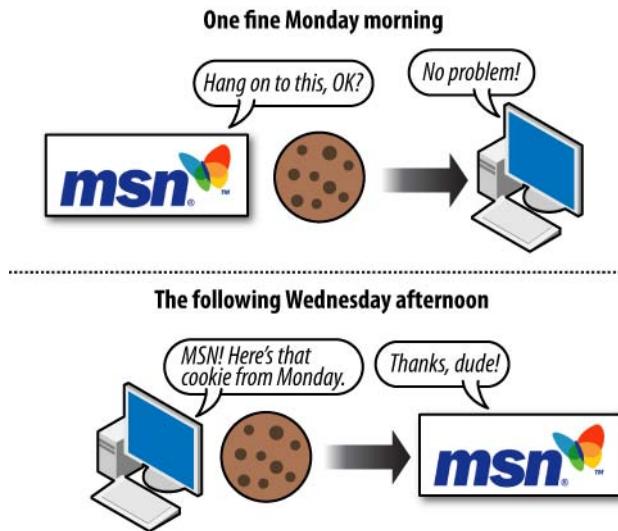


FIGURE 6-28. Using cookies to store data

this sense, they are virtually identical to cookies. Session IDs do come with an important caveat, though: they are frequently passed in the URL string, which can create serious problems for search engines (as every request produces a unique URL with duplicate content). A simple fix is to use the canonical tag (which we'll discuss in “Content Delivery and Search Spider Control” on page 238) to tell the search engines that you want them to ignore the session IDs.

N O T E

Any user has the ability to turn off cookies in his browser settings. This often makes web browsing considerably more difficult, and many sites will actually display a page saying that cookies are required to view or interact with their content. Cookies, persistent though they may be, are also deleted by users on a semiregular basis. For example, a 2007 comScore study found that 33% of web users deleted their cookies at least once per month.

How Do Search Engines Interpret Cookies and Session IDs?

They don't. Search engine spiders are not built to maintain or retain cookies or session IDs and act as browsers with this functionality shut off. However, unlike visitors whose browsers won't accept cookies, the crawlers can sometimes reach sequestered content by virtue of webmasters who want to specifically let them through. Many sites have pages that require cookies or sessions to be enabled but have special rules for search engine bots, permitting them to access the content as well. Although this is technically cloaking, there is a form of this known as First



FIGURE 6-29. How session IDs are used

Click Free that search engines generally allow (we will discuss this in more detail in “Content Delivery and Search Spider Control” on page 238).

Despite the occasional access engines are granted to cookie/session-restricted pages, the vast majority of cookie and session ID usage creates content, links, and pages that limit access. Web developers can leverage the power of concepts such as First Click Free to build more intelligent sites and pages that function in optimal ways for both humans and engines.

Why Would You Want to Use Cookies or Session IDs to Control Search Engine Access?

There are numerous potential tactics to leverage cookies and session IDs for search engine control. Here are many of the major strategies you can implement with these tools, but there are certainly limitless other possibilities:

Showing multiple navigation paths while controlling the flow of link juice

Visitors to a website often have multiple ways in which they'd like to view or access content. Your site may benefit from offering many paths to reaching content (by date, topic, tag, relationship, ratings, etc.), but expends PageRank or link juice that would be better optimized by focusing on a single, search-engine-friendly navigational structure. This is important because these varied sort orders may be seen as duplicate content.

You can require a cookie for users to access the alternative sort order versions of a page, and prevent the search engine from indexing multiple pages with the same content. One alternative solution to this is to use the canonical tag to tell the search engine that these alternative sort orders are really just the same content as the original page (we will discuss the canonical tag in “[Content Delivery and Search Spider Control](#)” on page 238).

Keep limited pieces of a page's content out of the engines' indexes

Many pages may contain content that you'd like to show to search engines and pieces you'd prefer appear only for human visitors. These could include ads, login-restricted information, links, or even rich media. Once again, showing noncookie users the plain version and cookie-accepting visitors the extended information can be invaluable. Note that this is often used in conjunction with a login, so only registered users can access the full content (such as on sites like Facebook and LinkedIn). For Yahoo! you can also use the robots-nocontent tag that allows you to specify portions of your page that Yahoo! should ignore (Google and Bing do not support the tag).

Grant access to pages requiring a login

As with snippets of content, there are often entire pages or sections of a site on which you'd like to restrict search engine access. This can be easy to accomplish with cookies/sessions, and it can even help to bring in search traffic that may convert to “registered-user” status. For example, if you had desirable content that you wished to restrict, you could create a page with a short snippet and an offer to continue reading upon registration, which would then allow access to that work at the same URL. We will discuss this more in “[Content Delivery and Search Spider Control](#)” on page 238.

Avoid duplicate content issues

One of the most promising areas for cookie/session use is to prohibit spiders from reaching multiple versions of the same content, while allowing visitors to get the version they prefer. As an example, at SEOMoz, logged-in users can see full blog entries on the blog home page, but search engines and nonregistered users will see only the snippets. This

prevents the content from being listed on multiple pages (the blog home page and the specific post pages), and provides a positive user experience for members.

Content Delivery and Search Spider Control

On occasion, it can be valuable to show search engines one version of content and show humans a different version. This is technically called cloaking, and the search engines' guidelines have near-universal policies restricting this. In practice, many websites, large and small, appear to use content delivery effectively and without being penalized by the search engines. However, use great care if you implement these techniques, and know the risks that you are taking.

Cloaking and Segmenting Content Delivery

Before we discuss the risks and potential benefits of cloaking-based practices, take a look at [Figure 6-30](#), which shows an illustration of how cloaking works.

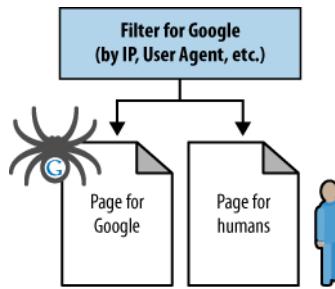


FIGURE 6-30. How cloaking works

Google's Matt Cutts, head of Google's webspam team, has made strong public statements indicating that all forms of cloaking (other than First Click Free) are subject to penalty. This was also largely backed by statements by Google's John Mueller in a May 2009 interview, which you can read at <http://www.stonetemple.com/articles/interview-john-mueller.shtml>.

Google makes its policy pretty clear in its Guidelines on Cloaking (<http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=66355>):

Serving up different results based on user agent may cause your site to be perceived as deceptive and removed from the Google index.

There are two critical pieces in the preceding quote: *may* and *user agent*. It is true that if you cloak in the wrong ways, with the wrong intent, Google and the other search engines *may* remove you from their index, and if you do it egregiously, they certainly *will*. But in some

cases, it may be the right thing to do, both from a user experience perspective and from an engine's perspective.

The key is intent: if the engines feel you are attempting to manipulate their rankings or results through cloaking, they may take adverse action against your site. If, however, the intent of your content delivery doesn't interfere with their goals, you're less likely to be subject to a penalty, as long as you don't violate important technical tenets (which we'll discuss shortly).

What follows are some examples of websites that perform some level of cloaking:

Google

Search for *google toolbar* or *google translate* or *adwords* or any number of Google properties and note how the URL you see in the search results and the one you land on almost never match. What's more, on many of these pages, whether you're logged in or not, you might see some content that is different from what's in the cache.

NYTimes.com

The interstitial ads, the request to log in/create an account after five clicks, and the archive inclusion are all showing different content to engines versus humans.

Wine.com

In addition to some redirection based on your path, there's the state overlay forcing you to select a shipping location prior to seeing any prices (or any pages). That's a form the engines don't have to fill out.

Yelp.com

Geotargeting through cookies based on location is a very popular form of local targeting that hundreds, if not thousands, of sites use.

Amazon.com

At SMX Advanced 2008 there was quite a lot of discussion about how Amazon does some cloaking (<http://www.naturalsearchblog.com/archives/2008/06/03/amazons-secret-to-dominating-serp-results/>). In addition, Amazon does lots of fun things with its *Buybox.amazon.com* subdomain and with the navigation paths and suggested products if your browser accepts cookies.

Trulia.com

Trulia was found to be doing some interesting redirects on partner pages and its own site (<http://www.bramblog.com/trulia-caught-cloaking-red-handed/>).

The message should be clear. Cloaking isn't always evil, it won't always get you banned, and you can do some pretty smart things with it. The key to all of this is your intent. If you are doing it for reasons that are not deceptive and that provide a positive experience for users and search engines, you might not run into problems. However, there is no guarantee of this, so use these types of techniques with great care, and know that you may still get penalized for it.

When to Show Different Content to Engines and Visitors

There are a few common causes for displaying content differently to different visitors, including search engines. Here are some of the most common ones:

Multivariate and A/B split testing

Testing landing pages for conversions requires that you show different content to different visitors to test performance. In these cases, it is best to display the content using JavaScript/cookies/sessions and give the search engines a single, canonical version of the page that doesn't change with every new spidering (though this won't necessarily hurt you). Google offers software called Google Website Optimizer to perform this function.

Content requiring registration and First Click Free

If you force registration (paid or free) on users to view specific content pieces, it is best to keep the URL the same for both logged-in and non-logged-in users and to show a snippet (one to two paragraphs is usually enough) to non-logged-in users and search engines. If you want to display the full content to search engines, you have the option to provide some rules for content delivery, such as showing the first one to two pages of content to a new visitor without requiring registration, and then requesting registration after that grace period. This keeps your intent more honest, and you can use cookies or sessions to restrict human visitors while showing the full pieces to the engines.

In this scenario, you might also opt to participate in a specific program from Google called First Click Free, wherein websites can expose "premium" or login-restricted content to Google's spiders, as long as users who click from the engine's results are given the ability to view that first article for free. Many prominent web publishers employ this tactic, including the popular site, Experts-Exchange.com.

To be specific, to implement First Click Free, the publisher must grant Googlebot (and presumably the other search engine spiders) access to all the content they want indexed, even if users normally have to log in to see the content. The user who visits the site will still need to log in, but the search engine spider will not have to do so. This will lead to the content showing up in the search engine results when applicable. However, if a user clicks on that search result, you must permit him to view the entire article (all pages of a given article if it is a multiple-page article). Once the user clicks to look at another article on your site, you can still require him to log in.

For more details, visit Google's First Click Free program page at <http://googlewebmastercentral.blogspot.com/2008/10/first-click-free-for-web-search.html>.

Navigation unspiderable to search engines

If your navigation is in Flash, JavaScript, a Java application, or another unspiderable format, you should consider showing search engines a version that has spiderable, crawlable content in HTML. Many sites do this simply with CSS layers, displaying a human-visible, search-invisible layer and a layer for the engines (and less capable browsers, such as mobile browsers). You can also employ the noscript tag for this purpose,

although it is generally riskier, as many spammers have applied `noscript` as a way to hide content. Adobe recently launched a portal on SEO and Flash and provides best practices that have been cleared by the engines to help make Flash content discoverable. Take care to make sure the content shown in the search-visible layer is substantially the same as it is in the human-visible layer.

Duplicate content

If a significant portion of a page's content is duplicated, you might consider restricting spider access to it by placing it in an iframe that's restricted by `robots.txt`. This ensures that you can show the engines the unique portion of your pages, while protecting against duplicate content problems. We will discuss this in more detail in the next section.

Different content for different users

At times you might target content uniquely to users from different geographies (such as different product offerings that are more popular in their area), with different screen resolutions (to make the content fit their screen size better), or who entered your site from different navigation points. In these instances, it is best to have a "default" version of content that's shown to users who don't exhibit these traits to show to search engines as well.

How to Display Different Content to Search Engines Versus Visitors

A variety of strategies exist to segment content delivery. The most basic is to serve content that is not meant for the engines in unspiderable formats (e.g., placing text in images, Flash files, plug-ins, etc.). You should not use these formats for the purpose of cloaking. You should use them only if they bring a substantial end-user benefit (such as an improved user experience). In such cases, you may want to show the search engines the same content in a search-spider-readable format. When you're trying to show the engines something you don't want visitors to see, you can use CSS formatting styles (preferably not `display:none`, as the engines may have filters to watch specifically for this), JavaScript, user-agent, cookie, or session-based delivery, or perhaps most effectively, IP delivery (showing content based on the visitor's IP address).

Be very wary when employing cloaking such as that we just described. The search engines expressly prohibit these practices in their guidelines, and though there is leeway based on intent and user experience (e.g., your site is using cloaking to improve the quality of the user's experience, not to game the search engines), the engines do take these tactics seriously and may penalize or ban sites that implement them inappropriately or with the intention of manipulation.

The `robots.txt` file

This file is located on the root level of your domain (e.g., <http://www.yourdomain.com/robots.txt>), and it is a highly versatile tool for controlling what the spiders are permitted to access on your site. You can use `robots.txt` to:

- Prevent crawlers from accessing nonpublic parts of your website
- Block search engines from accessing index scripts, utilities, or other types of code
- Avoid the indexation of duplicate content on a website, such as “print” versions of HTML pages, or various sort orders for product catalogs
- Auto-discover XML Sitemaps

The *robots.txt* file must reside in the root directory, and the filename must be entirely in lowercase (*robots.txt*, not *Robots.txt*, or other variations including uppercase letters). Any other name or location will not be seen as valid by the search engines. The file must also be entirely in text format (not in HTML format).

When you tell a search engine robot not to access a page, it prevents the crawler from accessing the page. [Figure 6-31](#) illustrates what happens when the search engine robot sees a direction in *robots.txt* not to crawl a web page.

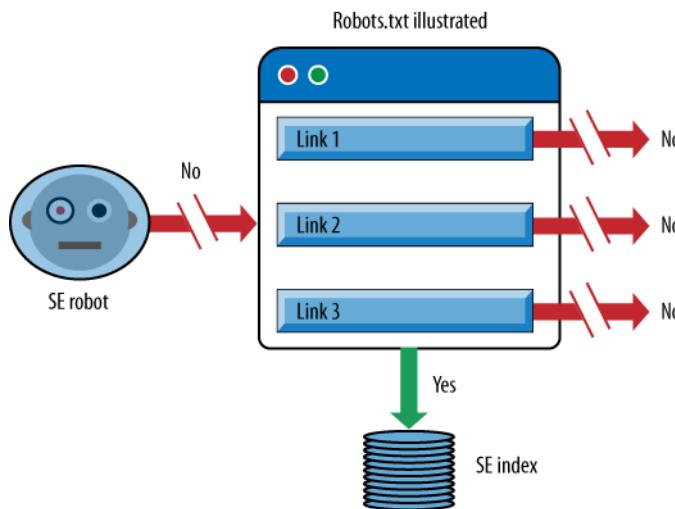


FIGURE 6-31. Impact of robots.txt

In essence, the page will not be crawled, so links on the page cannot pass link juice to other pages since the search engine does not see the links. However, the page can be in the search engine index. This can happen if other pages on the Web link to the page. Of course, the search engine will not have very much information on the page since it cannot read it, and will rely mainly on the anchor text and other signals from the pages linking to it to determine what the page may be about. Any resulting search listings end up being pretty sparse when you see them in the Google index, as shown in [Figure 6-32](#).



FIGURE 6-32. SERPs for pages that are listed in robots.txt

Figure 6-32 shows the results for the Google query `site:news.yahoo.com/topics/ inurl:page`. This is not a normal query that a user would enter, but you can see what the results look like. Only the URL is listed, and there is no description. This is because the spiders aren't permitted to read the page to get that data. In today's algorithms, these types of pages don't rank very high because their relevance scores tend to be quite low for any normal queries.

Google, Yahoo!, Bing, Ask, and nearly all of the legitimate crawlers on the Web will follow the instructions you set out in the `robots.txt` file. Commands in `robots.txt` are primarily used to prevent spiders from accessing pages and subfolders on a site, though they have other options as well. Note that subdomains require their own `robots.txt` files, as do files that reside on an https: server.

Syntax of the robots.txt file. The basic syntax of `robots.txt` is fairly simple. You specify a robot name, such as "googlebot", and then you specify an action. The robot is identified by user agent, and then the actions are specified on the lines that follow. Here are the major actions you can specify:

- **Disallow:** the pages you want to block the bots from accessing (as many disallow lines as needed)
- **Noindex:** the pages you want a search engine to block *and* not index (or de-index if previously indexed); this is unofficially supported by Google and unsupported by Yahoo! and Bing

Some other restrictions apply:

- Each User-Agent/Disallow group should be separated by a blank line; however, no blank lines should exist within a group (between the User-Agent line and the last Disallow).
- The hash symbol (#) may be used for comments within a *robots.txt* file, where everything after # on that line will be ignored. This may be used either for whole lines or for the end of lines.
- Directories and filenames are case-sensitive: “private”, “Private”, and “PRIVATE” are all uniquely different to search engines.

Here is an example of a *robots.txt* file:

```
User-agent: Googlebot
Disallow:

User-agent: msnbot
Disallow: /

# Block all robots from tmp and logs directories
User-agent: *
Disallow: /tmp/
Disallow: /logs      # for directories and files called logs
```

The preceding example will do the following:

- Allow “Googlebot” to go anywhere.
- Prevent “msnbot” from crawling any part of the site.
- Block all robots (other than Googlebot) from visiting the */tmp/* directory or directories or files called */logs* (e.g., */logs* or *logs.php*).

Notice that the behavior of Googlebot is not affected by instructions such as `Disallow: /`. Since Googlebot has its own instructions from *robots.txt*, it will ignore directives labeled as being for all robots (i.e., uses an asterisk).

One common problem that novice webmasters run into occurs when they have SSL installed so that their pages may be served via HTTP and HTTPS. A *robots.txt* file at `http://www.yourdomain.com/robots.txt` will not be interpreted by search engines as guiding their crawl behavior on `https://www.yourdomain.com`. To do this, you need to create an additional *robots.txt* file at `https://www.yourdomain.com/robots.txt`. So, if you want to allow crawling of all pages served from your HTTP server and prevent crawling of all pages from your HTTPS server, you would need to implement the following:

For HTTP:

```
User-agent: *
Disallow:
```

For HTTPS:

```
User-agent: *
Disallow: /
```

These are the most basic aspects of *robots.txt* files, but there are more advanced techniques as well. Some of these methods are supported by only some of the engines, as detailed in the list that follows:

Crawl delay

Crawl delay is supported by Yahoo!, Bing, and Ask. It instructs a crawler to wait the specified number of seconds between crawling pages. The goal with the directive is to reduce the load on the publisher's server:

```
User-agent: msnbot
Crawl-delay: 5
```

Pattern matching

Pattern matching appears to be usable by Google, Yahoo!, and Bing. The value of pattern matching is considerable. You can do some basic pattern matching using the asterisk wildcard character. Here is how you can use pattern matching to block access to all subdirectories that begin with *private* (e.g., */private1/*, */private2/*, */private3/*, etc.):

```
User-agent: Googlebot
Disallow: /private*/
```

You can match the end of the string using the dollar sign (\$). For example, to block URLs that end with *.asp*:

```
User-agent: Googlebot
Disallow: /*.asp$
```

You may wish to prevent the robots from accessing any URLs that contain parameters in them. To block access to all URLs that include a question mark (?), simply use the question mark:

```
User-agent: *
Disallow: /*?*
```

The pattern-matching capabilities of *robots.txt* are more limited than those of programming languages such as Perl, so the question mark does not have any special meaning and can be treated like any other character.

Allow directive

The *Allow* directive appears to be supported only by Google, Yahoo!, and Ask. It works the opposite of the *Disallow* directive and provides the ability to specifically call out directories or pages that may be crawled. When this is implemented it can partially override a previous *Disallow* directive. This may be beneficial after large sections of the site have been disallowed, or if the entire site itself has been disallowed.

Here is an example that allows Googlebot into only the *google* directory:

```
User-agent: Googlebot
Disallow: /
Allow: /google/
```

Noindex directive

This directive works in the same way as the `meta robots noindex` command (which we will discuss shortly) and tells the search engines to explicitly exclude a page from the index. Since a `Disallow` directive prevents crawling but not indexing, this can be a very useful feature to ensure that the pages don't show in search results. However, as of October 2009, only Google supports this directive in *robots.txt*.

Sitemaps

We discussed XML Sitemaps at the beginning of this chapter. You can use *robots.txt* to provide an autodiscovery mechanism for the spider to find the XML Sitemap file. The search engines can be told to find the file with one simple line in the *robots.txt* file:

```
Sitemap: sitemap_location
```

The `sitemap_location` should be the complete URL to the Sitemap, such as `http://www.yourdomain.com/sitemap.xml`. You can place this anywhere in your file.

For full instructions on how to apply *robots.txt*, see [Robots.txt.org](#). You may also find it valuable to use Dave Naylor's *robots.txt* generation tool to save time and heartache (<http://www.davidnaylor.co.uk/the-robotstxt-builder-a-new-tool.html>).

You should use great care when making changes to *robots.txt*. A simple typing error can, for example, suddenly tell the search engines to no longer crawl any part of your site. After updating your *robots.txt* file it is always a good idea to check it with the [Google Webmaster Tools Test Robots.txt tool](#).

The Rel="NoFollow" attribute

In 2005, the three major search engines (Yahoo!, Google, and Microsoft) all agreed to support an initiative intended to reduce the effectiveness of automated spam. Unlike the `meta robots` version of `NoFollow`, the new directive could be employed as an attribute within an `<a>` or link tag to indicate that the linking site "does not editorially vouch for the quality of the linked-to page." This enables a content creator to link to a web page without passing on any of the normal search engine benefits that typically accompany a link (things such as trust, anchor text, PageRank, etc.).

Originally, the intent was to enable blogs, forums, and other sites where user-generated links were offered to shut down the value of spammers who built crawlers that automatically created links. However, this has expanded as Google, in particular, recommends use of `NoFollow` on links that are paid for—as the search engine's preference is that only those links that are truly editorial and freely provided by publishers (without being compensated) should count toward bolstering a site's/page's rankings.

You can implement `NoFollow` using the following format:

```
<a href="http://www.google.com" rel="NoFollow">
```

Note that although you can use `NoFollow` to restrict the passing of link value between web pages, the search engines may still crawl through those links (despite the lack of semantic logic) and crawl the pages they link to. The search engines have provided contradictory input on this point. To summarize, `NoFollow` does not expressly forbid indexing or spidering, so if you link to your own pages with it, intending to keep those pages from being indexed or ranked, others may find them and link to them, and your original goal will be thwarted.

[Figure 6-33](#) shows how a search engine robot interprets a `NoFollow` attribute when it finds one associated with a link (Link 1 in this example).

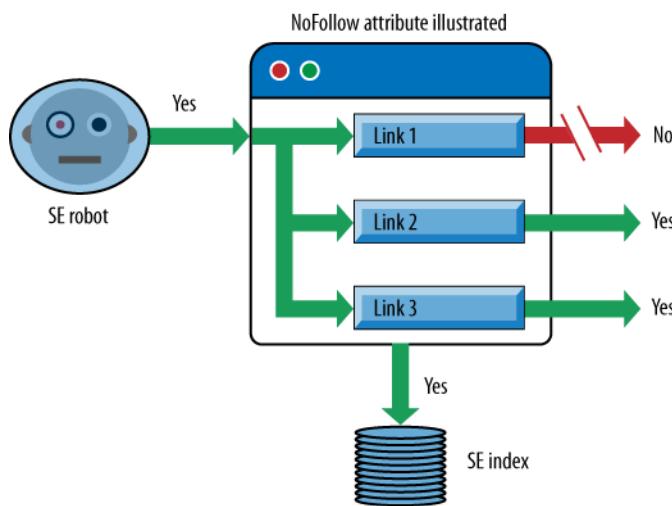


FIGURE 6-33. Impact of `NoFollow` attribute

The specific link with the `NoFollow` attribute is disabled from passing link juice. No other aspects of how the search engines deal with the page have been altered.

After the introduction of the `NoFollow` attribute, the notion of PageRank sculpting using `NoFollow` was a popular idea. The belief was that when you `NoFollow` a particular link, the link juice that would have been passed to that link was preserved and the search engines would reallocate it to the other links found on the page. As a result, many publishers implemented `NoFollow` links to lower value pages on their site (such as the About Us and Contact Us pages, or alternative sort order pages for product catalogs). In fact, data from [SEOMoz's Linkscape tool](#), published in March 2009, showed that at that time about 3% of all links on the Web were `NoFollowed`, and that 60% of those `NoFollows` were applied to internal links.

In June 2009, however, Google's Matt Cutts wrote a post that made it clear that the link juice associated with that `NoFollowed` link is discarded rather than reallocated (<http://www.mattcutts.com/blog/pagerank-sculpting/>). In theory, you can still use `NoFollow` however you want, but using it on internal links does not (at the time of this writing, according to Google) bring the type of benefit people have been looking for in the past. In fact, in certain scenarios it can be harmful.

The following example will illustrate the issue. If a publisher has a 500-page site, every page links to its About Us page, and all those links are `NoFollowed`, it will have cut off the link juice that would otherwise be sent to the About Us page. However, since that link juice is discarded, no benefit is brought to the rest of the site. Further, if the `NoFollows` are removed, the About Us page would pass at least some of that link juice back to the rest of the site through the links on the About Us page.

This is a great illustration of the ever-changing nature of SEO. Something that was a popular, effective tactic is now being viewed as ineffective. Some more aggressive publishers will continue to pursue link juice sculpting by using even more aggressive approaches, such as implementing links in encoded JavaScript or within iframes that have been disallowed in `robots.txt`, so that the search engines don't see them as links. Such aggressive tactics are probably not worth the trouble for most publishers.

The meta robots tag

The `meta robots` tag has three components: `cache`, `index`, and `follow`. The `cache` component instructs the engine about whether it can keep the page in the engine's public index, available via the "cached snapshot" link in the search results (see [Figure 6-34](#)).



FIGURE 6-34. Snapshot of cached element in the SERPs

The second, `index`, tells the engine whether the page is allowed to be crawled and stored in any capacity. A page marked `NoIndex` will thus be excluded entirely by the search engines. By

default, this value is `index`, telling the search engines, “Yes, please do crawl this page and include it in your index.” Thus, it is unnecessary to place this directive on each page. [Figure 6-35](#) shows what a search engine robot does once it sees a `NoIndex` tag on a web page.

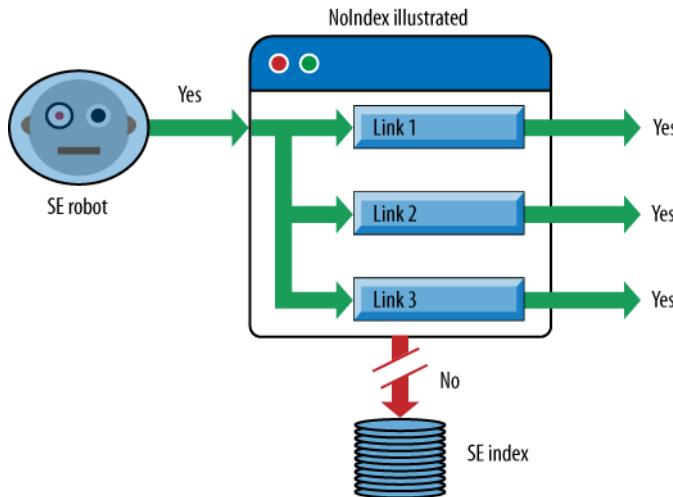


FIGURE 6-35. Impact of `NoIndex`

The page will still be crawled, and the page can still accumulate and pass link juice to other pages, but it will not appear in search indexes.

The final instruction available through the `meta robots` tag is `follow`. This command, like `index`, defaults to “yes, crawl the links on this page and pass link juice through them.” Applying `NoFollow` tells the engine that the links on that page should not pass link value or be crawled. By and large, it is unwise to use this directive as a way to prevent links from being crawled. Since human beings will still reach those pages and have the ability to link to them from other sites, `NoFollow` (in the `meta robots` tag) does little to restrict crawling or spider access. Its only application is to prevent link juice from spreading out, which has very limited application since the 2005 launch of the `rel="nofollow"` attribute (discussed earlier), which allows this directive to be placed on individual links.

[Figure 6-36](#) outlines the behavior of a search engine robot when it finds a `NoFollow` meta tag on a web page.

When you use the `NoFollow` meta tag on a page, the search engine will still crawl the page and place the page in its index. However, all links (both internal and external) on the page will be disabled from passing link juice to other pages.

One good application for `NoIndex` is to place this tag on HTML sitemap pages. These are pages designed as navigational aids for users and search engine spiders to enable them to efficiently

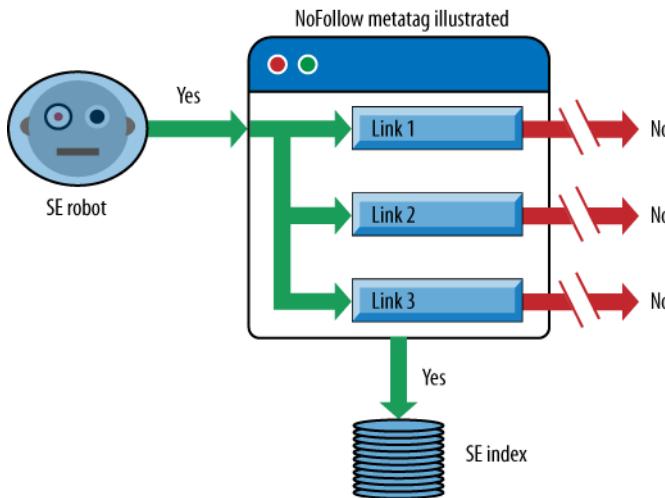


FIGURE 6-36. Impact of NoFollow meta tag

find the content on your site. However, on some sites these pages are unlikely to rank for anything of importance in the search engines, yet you still want them to pass link juice to the pages they link to. Putting `NoIndex` on these pages keeps these HTML sitemaps out of the index and removes that problem. Make sure you *do not* apply the `NoFollow` meta tag on the pages or the `NoFollow` attribute on the links on the pages, as these will prevent the pages from passing link juice.

The canonical tag

In February 2009, Google, Yahoo!, and Microsoft announced a new tag known as the canonical tag. This tag was a new construct designed explicitly for purposes of identifying and dealing with duplicate content. Implementation is very simple and looks like this:

```
<link rel="canonical" href="http://www.seomoz.org/blog" />
```

This tag is meant to tell Yahoo!, Bing, and Google that the page in question should be treated as though it were a copy of the URL <http://www.seomoz.org/blog> and that all of the link and content metrics the engines apply should technically flow back to that URL (see Figure 6-37).

The canonical URL tag attribute is similar in many ways to a 301 redirect from an SEO perspective. In essence, you're telling the engines that multiple pages should be considered as one (which a 301 does), without actually redirecting visitors to the new URL (often saving your development staff trouble). There are some differences, though:

- Whereas a 301 redirect points all traffic (bots and human visitors), the canonical URL tag is just for engines, meaning you can still separately track visitors to the unique URL versions.

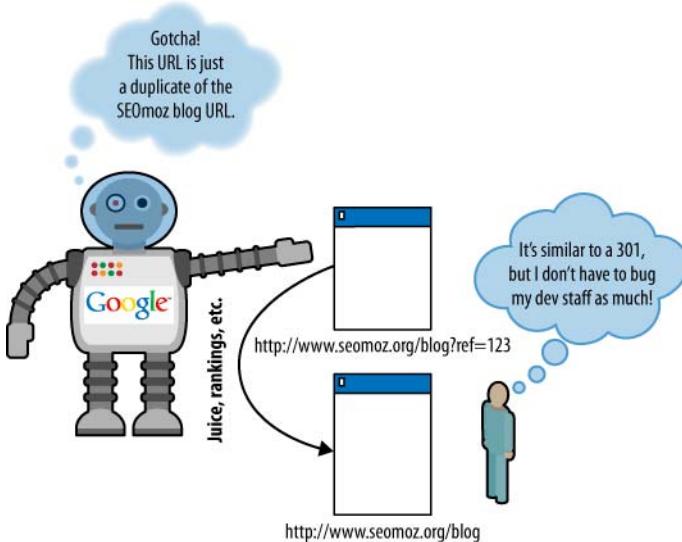


FIGURE 6-37. How search engines look at the canonical tag

- A 301 is a much stronger signal that multiple pages have a single, canonical source. Although the engines are certainly planning to support this new tag and trust the intent of site owners, there will be limitations. Content analysis and other algorithmic metrics will be applied to ensure that a site owner hasn't mistakenly or manipulatively applied the tag, and you can certainly expect to see mistaken use of the canonical tag, resulting in the engines maintaining those separate URLs in their indexes (meaning site owners would experience the same problems noted in “[Duplicate Content Issues](#)” on page 226).
- 301s carry cross-domain functionality, meaning you can redirect a page at Domain1.com to Domain2.com and carry over those search engine metrics. This is *not* the case with the canonical URL tag, which operates exclusively on a single root domain (it will carry over across subfolders and subdomains).

We will discuss some applications for this tag later in this chapter. In general practice, the best solution is to resolve the duplicate content problems at their core, and eliminate them if you can. This is because the canonical tag is not guaranteed to work. However, it is not always possible to resolve the issues by other means, and the canonical tag provides a very effective backup plan.

Blocking and cloaking by IP address range

You can customize entire IP addresses or ranges to block particular bots through server-side restrictions on IPs. Most of the major engines crawl from a limited number of IP ranges, making it possible to identify them and restrict access. This technique is, ironically, popular with

webmasters who mistakenly assume that search engine spiders are spammers attempting to steal their content, and thus block the IP ranges to restrict access and save bandwidth. Use caution when blocking bots, and make sure you're not restricting access to a spider that could bring benefits, either from search traffic or from link attribution.

Blocking and cloaking by user agent

At the server level, it is possible to detect user agents and restrict their access to pages or websites based on their declaration of identity. As an example, if a website detected a rogue bot, you might double-check its identity before allowing access. The search engines all use a similar protocol to verify their user agents via the Web: a reverse DNS lookup followed by a corresponding forward DNS→IP lookup. An example for Google would look like this:

```
> host 66.249.66.1
1.66.249.66.in-addr.arpa domain name pointer crawl-66-249-66-1.googlebot.com.

> host crawl-66-249-66-1.googlebot.com
crawl-66-249-66-1.googlebot.com has address 66.249.66.1
```

A reverse DNS lookup by itself may be insufficient, because a spoofing bot could set up reverse DNS to point to <xyz.googlebot.com> or any other address.

Using iframes

Sometimes there's a certain piece of content on a web page (or a persistent piece of content throughout a site) that you'd prefer search engines didn't see. As we discussed earlier in this chapter, clever use of iframes can come in handy, as [Figure 6-38](#) illustrates.

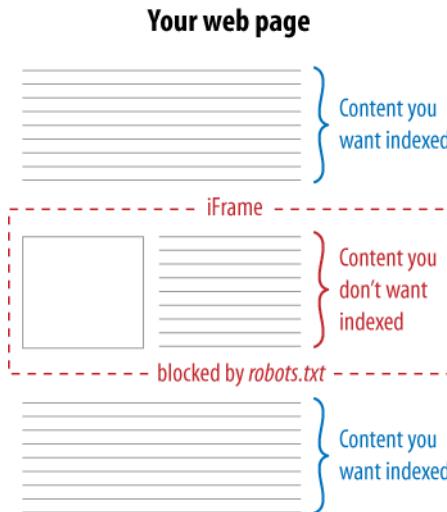


FIGURE 6-38. Using iframes to prevent indexing of content

The concept is simple: by using iframes, you can embed content from another URL onto any page of your choosing. By then blocking spider access to the iframe with *robots.txt*, you ensure that the search engines won't "see" this content on your page. Websites may do this for many reasons, including avoiding duplicate content problems, reducing the page size for search engines, or lowering the number of crawlable links on a page (to help control the flow of link juice).

Hiding text in images

As we discussed previously, the major search engines still have very little capacity to read text in images (and the processing power required makes for a severe barrier). Hiding content inside images isn't generally advisable, as it can be impractical for alternative devices (mobile, in particular) and inaccessible to others (such as screen readers).

Hiding text in Java applets

As with text in images, the content inside Java applets is not easily parsed by the search engines, though using them as a tool to hide text would certainly be a strange choice.

Forcing form submission

Search engines will not submit HTML forms in an attempt to access the information retrieved from a search or submission. Thus, if you keep content behind a forced-form submission and never link to it externally, your content will remain out of the engines (as [Figure 6-39](#) demonstrates).



FIGURE 6-39. Use of forms, which are unreadable by crawlers

The problem comes when content behind forms earns links outside your control, as when bloggers, journalists, or researchers decide to link to the pages in your archives without your knowledge. Thus, although form submission may keep the engines at bay, make sure that anything truly sensitive has additional protection (e.g., through *robots.txt* or meta robots).

Using login/password protection

Password protection of any kind will effectively prevent any search engines from accessing content, as will any form of human-verification requirements, such as CAPTCHAs (the boxes that request the copying of letter/number combinations to gain access). The major engines won't try to guess passwords or bypass these systems.

Removing URLs from a search engine's index

A secondary, post-indexing tactic, URL removal is possible at most of the major search engines through verification of your site and the use of the engines' tools. For example, Yahoo! allows you to remove URLs through its Site Explorer system (<http://help.yahoo.com/l/us/yahoo/search/siteexplorer/delete/siteexplorer-46.html>), and Google offers a similar service (<https://www.google.com/webmasters/tools/removals>) through Webmaster Central. Microsoft's Bing search engine may soon carry support for this as well.

Redirects

A redirect is used to indicate when content has moved from one location to another. For example, you may have some content at <http://www.yourdomain.com/old.html> and decide to restructure your site. As a result of this move, your content may move to <http://www.yourdomain.com/critical-keyword.html>.

Once a redirect is implemented users who go to the old versions of your pages (perhaps via a bookmark they kept for the page) will be sent to the new versions of those pages. Without the redirect, the user would get a Page Not Found (404) error. With the redirect, the web server tells the incoming user agent (whether a browser or a spider) to instead fetch the requested content from the new URL.

Why and When to Redirect

Redirects are also important for letting search engines know when you have moved content. After doing so, the search engines will continue to have the old URL in their index and return it in their search results. The solution to this is to implement a redirect. Here are some scenarios in which you may end up needing to implement redirects:

- You have old content that expires, so you remove it.
- You find that you have broken URLs that have links and traffic.

- You change your hosting company.
- You change your CMS.
- You want to implement a canonical redirect (redirect all pages on `http://yourdomain.com` to `http://www.yourdomain.com`).
- You change the URLs where your existing content can be found for any reason.

Not all of these scenarios require a redirect. For example, you can change hosting companies without impacting any of the URLs used to find content on your site, in which case no redirect is required. However, any scenario in which any of your URLs change is a scenario in which you need to implement redirects.

Good and Bad Redirects

It turns out that there are many ways to perform a redirect. Not all are created equal. The basic reason for this is that there are two major types of redirects that can be implemented, tied specifically to the HTTP status code returned by the web server to the browser. These are:

“301 moved permanently”

This status code tells the browser (or search engine crawler) that the resource has been permanently moved to another location, and there is no intent to ever bring it back.

“302 moved temporarily”

This status code tells the browser (or search engine crawler) that the resource has been temporarily moved to another location, and that the move should not be treated as permanent.

Both forms of redirect send a human or a search engine crawler to the new location, but the search engines interpret these two HTTP status codes in very different ways. When a crawler sees a 301 HTTP status code, it assumes it should pass the historical link juice (and any other metrics) from the old page to the new one. When a search engine crawler sees a 302 HTTP status code, it assumes it should not pass the historical link juice from the old page to the new one. In addition, the 301 redirect will lead the search engine to remove the old page from the index and replace it with the new one.

The preservation of historical link juice is very critical in the world of SEO. For example, imagine you had 1,000 links to `http://www.yourolddomain.com` and you decided to relocate everything to `http://www.yournewdomain.com`. If you used redirects that returned a 302 status code, you would be starting your link-building efforts from scratch again. In addition, the old version of the page may remain in the index and compete for search rankings in the search engines.

It should also be noted that there can be redirects that pass no status code, or the wrong status code, such as a 404 error (i.e., page not found) or a 200 OK (page loaded successfully). These are also problematic, and should be avoided. You want to definitively return a 301 HTTP status

code when you have performed a redirect whenever you make a permanent move to a page's location.

Methods for URL Redirecting and Rewriting

There are many possible ways to implement redirects. On Apache web servers (normally present on machines running Unix or Linux as the operating system), it is possible to implement redirects quite simply in a standard file called *.htaccess* using the `Redirect` and `RedirectMatch` directives (learn more about this file format at <http://httpd.apache.org/docs/2.2/howto/htaccess.html>). More advanced directives known as *rewrite rules* can be employed as well using the Apache module known as `mod_rewrite`, which we will discuss in a moment.

On web servers running Microsoft IIS (<http://www.iis.net/>), different methods are provided for implementing redirects. The basic method for doing these redirects is through the IIS console (you can read more about this at <http://www.mcanerin.com/EN/articles/301-redirect-IIS.asp>).

People with IIS servers can also make use of a text file with directives provided they use an ISAPI plug-in such as ISAPI_Rewrite (<http://www.isapirewrite.com/>), and this scripting language offers similar capabilities as that of Apache's `mod_rewrite` module.

Many programmers use other techniques for implementing redirects. This can be done directly in programming languages such as Perl, PHP, ASP, and JavaScript. The key thing that the programmer must do, if he implements redirects in this fashion, is to make sure the HTTP status code returned by the web server is a 301. You can check the header returned with the [Firefox plug-in Live HTTP Headers](#).

Another method that you can use to implement a redirect occurs at the page level, via the `meta refresh` tag, which looks something like this:

```
<meta http-equiv="refresh" content="5;url=http://www.yourdomain.com/newlocation.htm" />
```

The first parameter in the content section in the preceding statement, the number 5, indicates the number of seconds the web server should wait before redirecting the user to the indicated page. This gets used in scenarios where the publisher wants to display a page letting the user know that he is going to get redirected to a different page than the one he requested.

The problem is that most meta refreshes are treated as though they are a 302 redirect. The sole exception to this is if you specify a redirect delay of 0 seconds. You will have to give up your helpful page telling the user that you are moving him, but the search engines appear to treat this as though it were a 301 redirect (to be safe, the best practice is simply to use a 301 redirect if at all possible).

Mod_rewrite and ISAPI_Rewrite for URL rewriting and redirecting

There is much more to discuss on this topic than we can reasonably address in this book. The following description is intended only as an introduction to help orient more technical readers, including web developers and site webmasters, on how rewrites and redirects function. To skip

this technical discussion, proceed to “[Redirecting a Home Page Index File Without Looping](#)” on page 261.

Mod_rewrite for Apache and ISAPI_Rewrite for Microsoft IIS Server offer very powerful ways to rewrite your URLs. Here are some reasons for using these powerful tools:

- You have changed your URL structure on your site so that content has moved from one location to another. This can happen when you change your CMS, or change your site organization for any reason.
- You want to map your search-engine-unfriendly URLs into friendlier ones.

If you are running Apache as your web server, you would place directives known as *rewrite rules* within your *.htaccess* file or your Apache configuration file (e.g., *httpd.conf* or the site-specific config file in the *sites_conf* directory). Similarly, if you are running IIS Server, you’d use an ISAPI plug-in such as ISAPI_Rewrite and place rules in an *httpd.ini* config file.

Note that rules can differ slightly on ISAPI_Rewrite compared to mod_rewrite, and the following discussion focuses on mod_rewrite. Your *.htaccess* file would start with:

```
RewriteEngine on
RewriteBase /
```

You should omit the second line if you’re adding the rewrites to your server config file, since *RewriteBase* is supported only in *.htaccess*. We’re using *RewriteBase* here so that you won’t have to have *^/* at the beginning of all the rules, just *^* (we will discuss regular expressions in a moment).

After this step, the rewrite rules are implemented. Perhaps you want to have requests for product page URLs of the format *http://www.yourdomain.com/products/123* to display the content found at *http://www.yourdomain.com/get_product.php?id=123*, without the URL changing in the Location bar of the user’s browser and without you having to recode the *get_product.php* script. Of course, this doesn’t replace all occurrences of dynamic URLs within the links contained on all the site pages; that’s a separate issue. You can accomplish this first part with a single rewrite rule, like so:

```
RewriteRule ^products/([0-9]+)/?$ /get_product.php?id=$1 [L]
```

The preceding example tells the web server that all requests that come into the */product/* directory should be mapped into requests to */get_product.php*, while using the subfolder to */product/* as a parameter for the PHP script.

The *^* signifies the start of the URL following the domain, *\$* signifies the end of the URL, *[0-9]* signifies a numerical digit, and the *+* immediately following it means one or more occurrences of a digit. Similarly, the *?* immediately following the */* means zero or one occurrence of a slash character. The *()* puts whatever is wrapped within it into memory. You can then access what’s been stored in memory with *\$1* (i.e., what is in the first set of parentheses). Not surprisingly, if you included a second set of parentheses in the rule, you’d

access that with \$2, and so on. The [L] flag saves on server processing by telling the rewrite engine to stop if it matches on that rule. Otherwise, all the remaining rules will be run as well.

Here's a slightly more complex example, where URLs of the format `http://www.yourdomain.com/webapp/wcs/stores/servlet/ProductDisplay?`

`storeId=10001&catalogId=10001&langId=-1&categoryID=4&productID=123` would be rewritten to `http://www.yourdomain.com/4/123.htm`:

```
RewriteRule ^([^\/]+)([^\/]+\.\htm$ /webapp/wcs/stores/servlet/ProductDisplay?storeId=10001&catalogId=10001&langId=-1&categoryID=$1&productID=$2 [QSA,L]
```

The [^\/] signifies any character other than a slash. That's because, within square brackets, ^ is interpreted as *not*. The [QSA] flag is for when you don't want the query string dropped (like when you want a tracking parameter preserved).

To write good rewrite rules you will need to become a master of *pattern matching* (which is simply another way to describe the use of regular expressions). Here are some of the most important special characters and how the rewrite engine interprets them:

- * means 0 or more of the immediately preceding character.
- + means 1 or more of the immediately preceding character.
- ? means 0 or 1 occurrence of the immediately preceding character.
- ^ means the beginning of the string.
- \$ means the end of the string.
- . means any character (i.e., it acts as a wildcard).
- \ “escapes” the character that follows; for example, \. means the dot is not meant to be a wildcard, but an actual character.
- ^ inside // brackets means *not*; for example, [^\/] means *not slash*.

It is incredibly easy to make errors in regular expressions. Some of the common gotchas that lead to unintentional substring matches include:

- Using .* when you should be using .+ since .* can match on nothing.
- Not “escaping” with a backslash a special character that you don't want interpreted, as when you specify . instead of \. and you really meant the dot character rather than any character (thus, `default.htm` would match on `defaulthtm`, and `default\.htm` would match only on `default.htm`).
- Omitting ^ or \$ on the assumption that the start or end is implied (thus, `default.htm` would match on `mydefault.html` whereas `^default\htm$` would match only on `default.htm`).
- Using “greedy” expressions that will match on all occurrences rather than stopping at the first occurrence.

The easiest way to illustrate what we mean by greedy is to provide an example:

```
RewriteRule ^(.*)/?index\.html$ /$1/ [L,R=301]
```

This will redirect requests for `http://www.yourdomain.com/blah/index.html` to `http://www.yourdomain.com/blah//`. This is probably not what was intended. Why did this happen? Because `.*` will capture the slash character within it before the `/?` gets to see it. Thankfully, there's an easy fix. Simply use `/^` or `.*?` instead of `.*` to do your matching. For example, use `^(.*?)/?` instead of `^(.*)/?` or `[^/]+/[^\?]/` instead of `.*/.*`.

So, to correct the preceding rule you could use the following:

```
RewriteRule ^(.*)/?index\.html$ /$1/ [L,R=301]
```

When wouldn't you use the following?

```
RewriteRule ^([^\?]*)/?index\.html$ /$1/ [L,R=301]
```

This is more limited because it will match only on URLs with one directory. URLs containing multiple subdirectories, such as `http://www.yourdomain.com/store/cheese/swiss/wheel/index.html`, would not match.

As you might imagine, testing/debugging is a big part of URL rewriting. When debugging, the `RewriteLog` and `RewriteLogLevel` directives are your friends! Set the `RewriteLogLevel` to 4 or more to start seeing what the rewrite engine is up to when it interprets your rules.

By the way, the `[R=301]` flag in the last few examples—as you might guess—tells the rewrite engine to do a 301 redirect instead of a standard rewrite.

There's another handy directive to use in conjunction with `RewriteRule`, called `RewriteCond`. You would use `RewriteCond` if you are trying to match on something in the query string, the domain name, or other things not present between the domain name and the question mark in the URL (which is what `RewriteRule` looks at).

Note that neither `RewriteRule` nor `RewriteCond` can access what is in the anchor part of a URL—that is, whatever follows a `#`—because that is used internally by the browser and is not sent to the server as part of the request. The following `RewriteCond` example looks for a positive match on the hostname before it will allow the rewrite rule that follows to be executed:

```
RewriteCond %{HTTP_HOST} !^www\.\.yourdomain\.com\$ [NC]
RewriteRule ^(.*)\$ http://www.yourdomain.com/\$1 [L,R=301]
```

Note the exclamation point at the beginning of the regular expression. The rewrite engine interprets that as *not*.

For any hostname other than `http://www.yourdomain.com`, a 301 redirect is issued to the equivalent canonical URL on the www subdomain. The `[NC]` flag makes the rewrite condition case-insensitive. Where is the `[QSA]` flag so that the query string is preserved, you might ask? It is not needed when redirecting; it is implied.

If you don't want a query string retained on a rewrite rule with a redirect, put a question mark at the end of the destination URL in the rule, like so:

```
RewriteCond %{HTTP_HOST} !^www\.\.yourdomain\.com$ [NC]
RewriteRule ^(.*)$ http://www.yourdomain.com/$1? [L,R=301]
```

Why not use `^yourdomain\.com$` instead? Consider:

```
RewriteCond %{HTTP_HOST} ^yourdomain\.com$ [NC]
RewriteRule ^(.*)$ http://www.yourdomain.com/$1? [L,R=301]
```

That would not have matched on typo domains, such as “`yourdoamin.com`”, that the DNS server and virtual host would be set to respond to (assuming that misspelling was a domain you registered and owned).

Under what circumstances might you want to omit the query string from the redirected URL, as we did in the preceding two examples? When a session ID or a tracking parameter (such as `source=banner_ad1`) needs to be dropped. Retaining a tracking parameter after the redirect is not only unnecessary (because the original URL with the source code appended would have been recorded in your access logfiles as it was being accessed); it is also undesirable from a canonicalization standpoint. What if you wanted to drop the tracking parameter from the redirected URL, but retain the other parameters in the query string? Here’s how you’d do it for static URLs:

```
RewriteCond %{QUERY_STRING} ^source=[a-zA-Z0-9]*$
RewriteRule ^(.*)$ /$1? [L,R=301]
```

And for dynamic URLs:

```
RewriteCond %{QUERY_STRING} ^(.+)&source=[a-zA-Z0-9]+(&?.*)$
RewriteRule ^(.*)$ /$1?%1%2 [L,R=301]
```

Need to do some fancy stuff with cookies before redirecting the user? Invoke a script that cookies the user and then 301s him to the canonical URL:

```
RewriteCond %{QUERY_STRING} ^source=([a-zA-Z0-9]*)$
RewriteRule ^(.*)$ /cookiefirst.php?source=%1&dest=$1 [L]
```

Note the lack of a `[R=301]` flag in the preceding code. That’s on purpose. There’s no need to expose this script to the user. Use a rewrite and let the script itself send the 301 after it has done its work.

Other canonicalization issues worth correcting with rewrite rules and the `[R=301]` flag include when the engines index online catalog pages under HTTPS URLs, and URLs missing a trailing slash that should be there. First, the HTTPS fix:

```
# redirect online catalog pages in the /catalog/ directory if HTTPS
RewriteCond %{HTTPS} on
RewriteRule ^catalog/(.*) http://www.yourdomain.com/catalog/$1 [L,R=301]
```

Note that if your secure server is separate from your main server, you can skip the `RewriteCond` line.

Now to append the trailing slash:

```
RewriteRule ^(.*/[^/])$ /$1/ [L,R=301]
```

After completing a URL rewriting project to migrate from dynamic URLs to static, you'll want to phase out the dynamic URLs not just by replacing all occurrences of the legacy URLs on your site, but also by 301-redirecting the legacy dynamic URLs to their static equivalents. That way, any inbound links pointing to the retired URLs will end up leading both spiders and humans to the correct new URL—thus ensuring that the new URLs are the ones that are indexed, blogged about, linked to, and bookmarked, and the old URLs will be removed from the index as well. Generally, here's how you'd accomplish that:

```
RewriteCond %{QUERY_STRING} id=([0-9]+)
RewriteRule ^get_product\.php$ /products/%1.html? [L,R=301]
```

However, you'll get an infinite loop of recursive redirects if you're not careful. One quick-and-dirty way to avoid that situation is to add a nonsense parameter to the destination URL for the rewrite and ensure that this nonsense parameter isn't present before doing the redirect.

Specifically:

```
RewriteCond %{QUERY_STRING} id=([0-9]+)
RewriteCond %{QUERY_STRING} !blah=blah
RewriteRule ^get_product\.php$ /products/%1.html? [L,R=301]
RewriteRule ^products/([0-9]+)/?$ /get_product.php?id=$1&blah=blah [L]
```

Notice that the example used two `RewriteCond` lines, stacked on top of each other. All redirect conditions listed together in the same block will be “ANDed” together. If you wanted the conditions to be “ORed”, it would require the use of the `[OR]` flag.

Redirecting a Home Page Index File Without Looping

Many websites link to their own home page in a form similar to `http://www.yourdomain.com/index.html`. The problem with that is that most incoming links to the site's home page specify `http://www.yourdomain.com`, thus dividing the link juice into the site. Once a publisher realizes this, they will want to fix their internal links and then 301 redirect `http://www.yourdomain.com/index.html` to `http://www.yourdomain.com/`, but there will be problems with recursive redirects that develop if this is not done correctly.

When someone comes to your website by typing in `http://www.yourdomain.com`, the DNS system of the Internet helps the browser locate the web server for your website. How, then, does the web server decide what to show to the browser? It turns out that it does this by loading a file from the hard drive of the web server for your website.

When no file is specified (i.e., as in the preceding example, only the domain name is specified), the web server loads a file that is known as the default file. This is often a file with a name such as `index.html`, `index.htm`, `index.shtml`, `index.php`, or `default.asp`.

The filename can actually be anything, but most web servers default to one type of filename or another. Where the problem comes in is that many CMSs will expose both forms of your home page, both `http://www.yourdomain.com` and `http://www.yourdomain.com/index.php`.

Perhaps all the pages on the site link only to <http://www.yourdomain.com/index.php>, but given human nature, most of the links to your home page that third parties give you will most likely point at <http://www.yourdomain.com/>. This can create a duplicate content problem if the search engine now sees two versions of your home page and thinks they are separate, but duplicate, documents. Google is pretty smart at figuring out this particular issue, but it is best to not rely on that.

Since you learned how to do 301 redirects, you might conclude that the solution is to 301-redirect <http://www.yourdomain.com/index.php> to <http://www.yourdomain.com/>. Sounds good, right? Unfortunately, there is a big problem with this.

What happens is the server sees the request for <http://www.yourdomain.com/index.php> and then sees that it is supposed to 301-redirect that to <http://www.yourdomain.com/>, so it does. But when it loads <http://www.yourdomain.com/> it retrieves the default filename (*index.php*) and proceeds to load <http://www.yourdomain.com/index.php>. Then it sees that you want to redirect that to <http://www.yourdomain.com/>, and it creates an infinite loop.

The default document redirect solution

The solution that follows is specific to the preceding *index.php* example. You will need to plug in the appropriate default filename for your own web server.

1. Copy the contents of *index.php* to another file. For this example, we'll be using *sitehome.php*.
2. Create an Apache `DirectoryIndex` directive for your document root. Set it to *sitehome.php*. Do not set the directive on a serverwide level; otherwise, it may cause problems with other folders that still need to use *index.php* as a directory index.
3. Put this in an *.htaccess* file in your document root: `DirectoryIndex sitehome.php`. Or, if you aren't using per-directory context files, put this in your *httpd.conf*:

```
<Directory /your/document/root/examplesite.com/>
    DirectoryIndex sitehome.php
</Directory>
```

4. Clear out the contents of your original *index.php* file. Insert this line of code:

```
<? header("Location: http://www.example.com"); ?>
```

This sets it up so that *index.php* is not a directory index file (i.e., the default filename). It forces *sitehome.php* to be read when someone types in the canonical URL (<http://www.yourdomain.com>). Any requests to *index.php* from old links can now be 301-redirected while avoiding an infinite loop.

If you are using a CMS, you also need to make sure when you are done with this that all the internal links now go to the canonical URL, <http://www.yourdomain.com>. If for any reason the CMS started to point to <http://www.yourdomain.com/sitehome.php> the loop problem would return, forcing you to go through this entire process again.

Content Management System (CMS) Issues

When looking to publish a new site, many publishers may wonder whether they need to use a CMS, and if so, how to ensure that it is SEO-friendly.

It is first essential to determine whether a CMS is necessary before embarking on a web development project. You can use the flowchart in [Figure 6-40](#) to help guide you through the process.

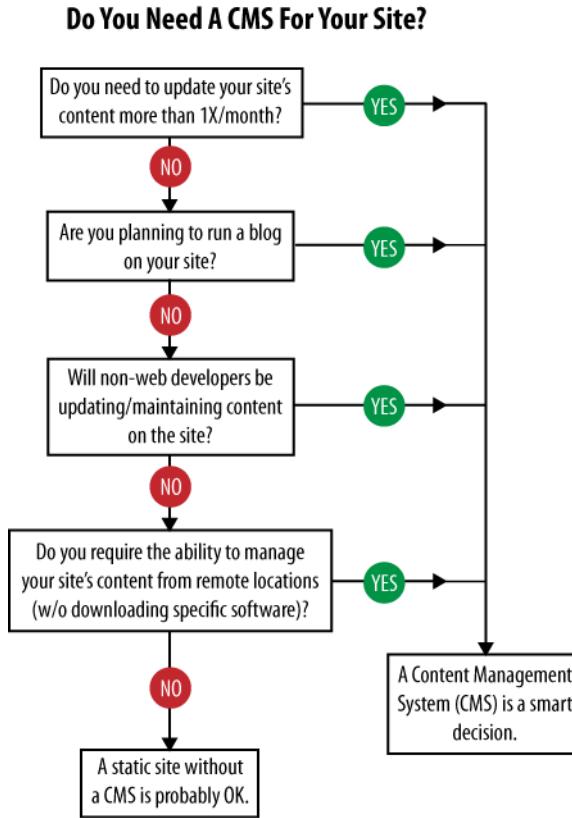


FIGURE 6-40. Flowchart to determine whether you need a CMS

Due to the inexpensiveness of customizable, free platforms such as [Drupal](#), [Joomla](#), [WordPress](#), and [Mambo](#), it is increasingly rare for a publisher to develop a static site, even when a CMS isn't required.

The next step involves understanding how to ensure that a CMS will be search-engine-friendly. Here is a list of basic SEO issues that frequently plague CMSs (both prebuilt and custom-made). By dealing with these, you will ensure a relatively smooth platform for content delivery.

Title tag customization and rules

A search-engine-friendly CMS must allow for title tags to not only be customized on a page-specific level, but also enable rules for particular sections of a website. For example, if the title tag always has to start with your site name followed by a colon followed by your article title, you're sunk—at least as far as your SEO is concerned. You should be able to revise the formulas used to generate the title tags across your site to make them more search-optimal.

Static, keyword-rich URLs

URLs have historically been the most problematic SEO issue for CMS platforms.

Nowadays, search-friendly CMSs should feature custom URL creation. In WordPress, a custom URL is referred to as a post slug. [Figure 6-41](#) is an example from SEOmoz's custom-built CMS.

Notice how the first line allows you to create the title of the post, and the second enables manual sculpting of the URL structure (and an automatic Generate button if you prefer to simply use the post title).



FIGURE 6-41. Example of custom URL creation

Meta tag customization

Being able to implement custom meta descriptions and meta robots tags is critical. Enabling editorial control is essential for a good CMS.

Enabling custom HTML tags

A good CMS has to offer extra functionality on HTML tags for things such as `NoFollow` on links, or `<hx>` tags for headlines and subheadlines. These can be built-in features accessible through menu options, or the CMS can simply allow a manual editing of HTML in the text editor window when required. Having no H1 tags on a given page is not desirable. Too many H1 tags on the page are not desirable. Low-value content (such as the publication date) marked up as an H1 is not desirable. The article title is typically the best content to have wrapped in an H1.

Internal anchor text flexibility

To be “optimized” rather than simply search-friendly, customizing the anchor text on internal links is critical. Rather than simply making all links in a site’s architecture the page’s title, a great CMS should be flexible enough to handle custom input from the administrators as to the anchor text of category-level or global navigation links.

Intelligent categorization structure

Another problem is poor category structure. When designing an information architecture for a website, you should not place limits on how pages are accessible due to the CMS's inflexibility. CMSs that offer customizable navigation panels will be the most successful in this respect.

Pagination controls

Pagination can be the bane of a website's search rankings, so controlling it through inclusion of more items per page, more contextually relevant anchor text (e.g., not "next," "prev," and page numbers), and careful use of `NoFollow` and meta `NoIndex` tags will make your important content get more link juice and crawl attention.

301-redirect functionality

Many CMSs sadly lack this critical feature, disallowing the proper redirection of content when necessary; 301s are valuable for expired content, for pages that have a newer version, and for dodging keyword cannibalization issues similar to those we discussed earlier in this chapter.

XML/RSS pinging

Although it is primarily useful for blogs, any content, from articles to products to press releases, can be issued in a feed, and by utilizing quick, accurate pinging of the major feed services, you limit some of your exposure to duplicate content spammers who pick up your feeds and ping the major services quickly in the hopes of beating you to the punch.

Image-handling and alt attributes

Alt attributes are a clear must-have from an SEO perspective, serving as the "anchor text" when an image is used as a link (note that text links are much better than images with alt attributes for links, but if you must use image links you do want to have the alt attribute implemented), and providing relevant, indexable content for the search engines. Images in a CMS's navigational elements should preferably use CSS image replacement rather than mere alt attributes.

CSS exceptions

The application of CSS styles in a proper CMS should allow for manual exceptions so that a user can modify how a strong headline or list element appears visually. If the CMS does not offer this, writers may opt out of using proper semantic markup for presentation purposes, which would not be a good thing.

Static caching options

Many CMSs currently offer caching options, which are a particular boon if a page is receiving a high level of traffic from social media portals or news sites. A bulky CMS often makes dozens of extraneous database connections, which can overwhelm a server if caching is not in place, killing potential inbound links and media attention.

URLs free of tracking parameters and session IDs

Sticking session or tracking information such as the user's click path into the URL is deadly for SEO. It usually leads to incomplete indexation and duplicate content issues.

Customizable URL structure

If the default URL structure of the CMS doesn't suit your needs, you should be able to change it. For example, if you don't want */archives/* in the URLs of all your archived articles, you should be able to remove it. Or if you want to reference the article name instead of the article's database ID in the URL, you should be able to do it.

301 redirects to a canonical URL

Duplicate content is the bane of many a dynamic website owner. Automatic handling of this by the CMS through the use of 301 redirects is a must.

Static-looking URLs

The most palatable URLs to spiders are the ones that look like they lead to static pages—no query strings in the URL.

Keywords in URLs

Keywords in your URLs can help your rankings.

RSS feeds

The CMS should auto-create RSS feeds to help your site rank in Google Blog Search and other feed engines.

Tagging and tag clouds

This Web 2.0 feature is powerful for SEO, thanks in large part to the keyword-rich text links.

Multilevel categorization structure

It is awfully limiting to your site structure and internal hierarchical linking structure to have a CMS that doesn't allow you to nest subcategories into categories, sub-subcategories into subcategories, and so on.

Paraphrasable excerpts

Duplicate content issues are exacerbated on dynamic sites such as blogs when the same content is displayed on Permalink pages, category pages, archives-by-date pages, tag pages, and the home page. Crafting unique content for the excerpt and having that content display on all locations except for the Permalink page will help strengthen your Permalink page as unique content.

Breadcrumb navigation

Breadcrumb (drill-down) navigation is great for SEO because it reinforces your internal hierarchical linking structure with keyword-rich text links.

Meta NoIndex tags for low-value pages

Even if you use **NoFollow** attributes in links to these pages, other people may still link to them, which carries a risk of ranking those pages above some of your more valuable content.

Keyword-rich intro copy on category-level pages

Keyword-rich introductory copy helps set a stable keyword theme for the page, rather than relying on the latest article or blog post to be the most prominent text on the page.

NoFollow links in comments

If you allow visitors to post comments and do not **NoFollow** the links, your site will be a spam magnet. Heck, you'll probably be a spam magnet anyway, but you won't risk losing PageRank to spammers if you use **NoFollow** attributes.

Customizable anchor text on navigational links

"Contact," "About Us," "Read More," "Full Article," and so on make for lousy anchor text—at least from an SEO standpoint. Hopefully, your CMS allows you to improve such links to make the anchor text more keyword-rich.

XML Sitemap generator

Having your CMS generate your XML Sitemap can save a lot of hassle, as opposed to trying to generate one with a third-party tool.

XHTML validation

Although HTML validation is not a ranking signal, it is desirable to have the CMS automatically check for malformed HTML, as search engines may end up seeing a page differently from how it renders on the screen and accidentally consider navigation to be part of the content or vice versa.

Pingbacks, trackbacks, comments, and antispam mechanisms

The problem with comments/trackbacks/pingbacks is that they are vectors for spam, so if you have one or more of these features enabled, you will be spammed. Therefore, effective spam prevention in the form of Akismet, Mollom, or Defensio is a must.

If you want more information on picking a quality CMS, some great web resources are already out there, among them <http://www.opensourcecms.com> and <http://www.cmsmatrix.org>, to help manage this task.

Selecting a CMS

There are many factors to consider when choosing an existing CMS. Many CMSs are free, but some of them are proprietary with a license cost per site. The majority of CMSs were not designed with security, stability, search friendliness, and scalability in mind, though in recent years a few vendors have developed excellent CMSs that have search friendliness as their primary focus. Many were developed to fit a certain market niche, but can be expanded to fit other purposes. Some are no longer maintained. Many are supported and developed primarily by hobbyists who don't particularly care if you're having trouble getting them installed and configured. Some are even intentionally made to be difficult to install and configure so that you'll be encouraged to pay the developers a consulting fee to do it all for you.

Once your CMS is up and running, ease of administration is a consideration. For instance, the CMS MODx would be very tough for a newbie to administer compared to [Blogger.com](#). If you just want to get content on the Web, Blogger will do about 90% of the most important SEO features, and any nontechnical user can use it—which makes it really attractive compared to MODx.

Selecting a CMS is an important process. If you make the wrong choice, you will doom your site to failure. Like most software, CMSs are a moving target—what's missing today may be a new feature tomorrow. In addition, just because a feature exists doesn't mean it is the default option, so in many instances the desired functionality will need to be enabled and possibly customized to work to your specifications.

Third-Party CMS Add-ons

Many CMS platforms offer many third-party plug-ins or add-ons that extend the core functionality of the CMS. In the WordPress plug-in directory alone there are nearly 6,000 plug-ins. Plug-ins provide a simple way to add new SEO features and functionality, making the CMS much more flexible and future-proof. It is particularly helpful when there is an active community developing plug-ins. An active community also comes in very handy in providing free technical support when things go wrong; and when bugs and security vulnerabilities crop up, it is important to have an active developer base to solve those issues quickly.

CMS add-ons can also come in the form of independent software installed on your web server or hosted services. Such tools come in many forms, discussion forums, customer reviews, and user polls. Discussion forums, for example, come in two forms, with bbPress, which is installed software and is optimized for search, and vbulletin, which is a hosted solution and therefore more difficult to optimize for search.

The problem with hosted solutions is that you are helping to build the service providers' link authority and not your own, and you have much less control over optimizing the content. Some hosted solutions can pose even bigger problems if the content and functionality are embedded into your site with JavaScript. Examples of this are leading customer review solutions such as BazaarVoice and PowerReviews.

One novel solution to the JavaScript problem is to execute the JavaScript, extract the content from its encrypted form, and present it in plain-text format so that the search engines can see it. This does require programming expertise to execute, but one tool that allows you to do this in a more automated fashion is Netconcepts' [GravityStream](#).

Optimizing Flash

As we referenced several times earlier in this chapter, Flash is popular on the Web, but each presents challenges to the search engines in terms of indexing the related content. This creates a gap between the user experience with a site and what the search engines can find on that site.

Flash

It used to be that search engines did not index Flash content at all. In June 2008, Google announced that it was offering improved indexing of this content (<http://googlewebmastercentral>

[.blogspot.com/2008/06/improved-flash-indexing.html](http://blogspot.com/2008/06/improved-flash-indexing.html)). This announcement indicates that Google can index text content and find and follow links within Flash files. However, Google still cannot tell what is contained in images within the Flash file. Here are some reasons why Flash is still not fully SEO-friendly:

*Different content is **not** on different URLs*

This is the same problem you encounter with AJAX-based pages. You could have unique frames, movies within movies, and so on that appear to be completely unique portions of the Flash site, yet there's often no way to link to these individual elements.

The breakdown of text is not clean

Google can index the output files in the SWF file to see words and phrases, but in Flash, a lot of your text is not inside clean `<h1>` or `<p>` tags; it is jumbled up into half-phrases for graphical effects and will often be output in the incorrect order. Worse still are text effects that often require “breaking” words apart into individual letters to animate them.

Flash gets embedded

A lot of Flash content is only linked to by other Flash content wrapped inside shell Flash pages. This line of links, where no other internal or external URLs are referencing the interior content, means some very low PageRank/link juice documents. Even if they manage to stay in the main index, they probably won't rank for anything.

Flash doesn't earn external links like HTML

An all-Flash site might get a large number of links to the home page, but interior pages almost always suffer. For embeddable Flash content, it is the HTML host page earning those links when they do come.

SEO basics are often missing

Anchor text, headlines, bold/strong text, `img alt` attributes, and even title tags are not simple elements to properly include in Flash. Developing Flash with SEO in mind is just more difficult than doing it in HTML. In addition, it is not part of the cultural lexicon of the Flash development world.

A lot of Flash isn't even crawlable

Google has indicated that it doesn't execute external JavaScript calls (which many Flash-based sites use) or index the content from external files called by Flash (which, again, a lot of Flash sites rely on). These limitations could severely impact what a visitor can see versus what Googlebot can index.

Note that it used to be that you could not test the crawlability of Flash, but the Adobe Search Engine SDK does allow you to get an idea as to how the search engines will see your Flash file.

Flash Coding Best Practices

If Flash is a requirement for whatever reason, there are best practices you can implement to make your site more accessible to search engine spiders. What follows are some guidelines on how to obtain the best possible results.

Flash meta tags

Beginning with Adobe/Macromedia Flash version 8, there has been support for the addition of title and description meta tags to any .swf file. Not all search engines are able to read these tags yet, but it is likely that they will soon. Get into the habit of adding accurate, keyword-rich title tags and meta tags to files now so that as search engines begin accessing them, your existing .swf files will already have them in place.

Adobe Flash search engine SDK

Flash developers may find the SDK useful for server-based text and link extraction and conversion purposes, or for client-side testing of their Flash content against the basic Adobe (formerly Macromedia) Flash Search Engine SDK code.

Tests have shown that Google and other major search engines now extract some textual content from Flash .swf files. It is unknown whether Google and others have implemented Adobe's specific Search Engine SDK technology into their spiders, or whether they are using some other code to extract the textual content. Again, tests suggest that what Google is parsing from a given .swf file is very close to what can be extracted manually using the Search Engine SDK.

The primary application of Adobe's Search Engine SDK is in the desktop testing of .swf files to see what search engines are extracting from a given file. The program cannot extract files directly from the Web; the .swf file must be saved to a local hard drive. The program is DOS-based and must be run in the DOS Command Prompt using DOS commands.

By running a .swf file through the Flash SDK swf2html program during development, the textual assets of the file can be edited or augmented to address the best possible SEO practices—honing in primarily on keywords and phrases along with high-quality links. Because of the nature of the Flash program and the way in which it deals with both text and animation, it is challenging to get exacting, quality SEO results. The goal is to create the best possible SEO results within the limitations of the Flash program and the individual Flash animation rather than to attempt the creation of an all-encompassing SEO campaign. Extracted content from Flash should be seen as one tool among many in a larger SEO campaign.

Internal Flash coding

There are several things to keep in mind when preparing Flash files for SEO:

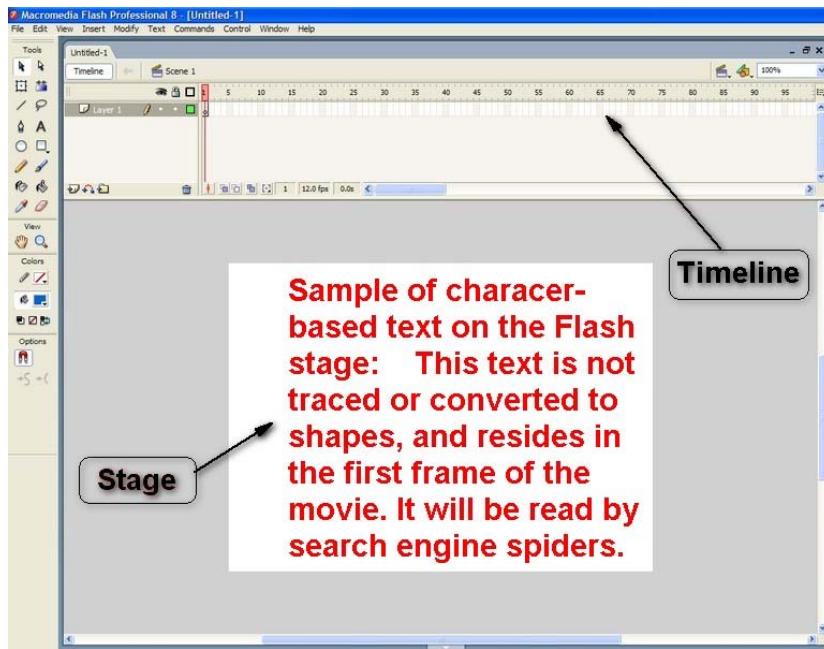


FIGURE 6-42. Example of spider-readable text inside a Flash program

- Search engines currently do not read traced text (using the `trace()` function) or text that has been transformed into a shape in Flash (as opposed to actual characters). Only character-based text that is active in the Flash stage will be read (see Figure 6-42).
- Animated or affected text often creates duplicate content. Static text in Flash movies is not read as duplicate instances that “tweening” and other effects can create. Use static text, especially with important content, so that search engines do not perceive the output as spam (see Figure 6-43).
- Search engine spiders do not see dynamically loaded content (text added from an external source, such as an XML file).
- The font size of text does not affect search engines; they read any size font.
- Special characters such as `<`, `>`, `&`, and `"` are converted to HTML character references (`<`, `>`, `&`, and `"`) and should be avoided.
- Search engines find and extract all URLs stored within the `getURL()` command.
- Search engines have the ability to follow links in Flash, though it is an “iffy” proposition at best. They will not, however, follow links to other Flash `.swf` files. (This is different from loading child `.swf` files into a parent `.swf` file.) Therefore, links in Flash should always point to HTML pages, not other `.swf` files.

```

<p>Fast Homework Help</p>
<p>Fast Homework Help</p>
<p>Reliable Report Info</p>
<p>Reliable Report Info</p>
<p>Math Made Easy</p>
<p>Math Made Easy</p>
<p>Brain Games + More</p>
<p>Brain Games + More</p>
<p>SCHOOL TOOLS</p>
<p>SCHOOL TOOLS</p>
<p>What Is Cosmeo?</p>
<p>What Is Cosmeo?</p>
<p>Why It Works</p>
<p>Why It Works</p>
<p>Take A Tour</p>
<p>Take A Tour</p>
<p>Get Started</p>
<p>Get Started</p>

```

FIGURE 6-43. Animated text results in Flash source; can be seen as duplicate content

SWFObject and NoScript tags

Because “alternative content” workarounds for SEO of Flash files have been historically abused by spammers, it is challenging to recommend these tactics to optimize your Flash files without a critical disclaimer.

Both the `SWFObject` and `NoScript` methods were originally designed to be legitimate, graceful degradation tactics readily accepted by the search engines as a way to accommodate older browsers or people with special needs. But many unscrupulous sites have used the code to trick search engine spiders. In other words, they are used in such a way that browsers display one thing to users, but display something completely different to search engine spiders. All of the major search engines disapprove of such tactics.

Websites using such methods today are often penalized or removed from search engine indexes altogether. This makes graceful degradation risky on some level, but if the methods are used clearly within the boundaries for which they were intended, getting penalized or banned is highly unlikely.

Intent is an essential element search engines take into consideration. If your intent is to provide *all* users with a positive experience while visiting your site, you should be fine. If your intent is to game the search engines, all it takes is one online rival to report your site for spam to incur the wrath of the search engines.

Google and other search engines do not algorithmically ban sites for using `SWFObject` and `NoScript` tags; it usually requires human intervention to evoke a penalty or outright ban.

SWFObject. With regard to Flash optimization, **SWFObject** is the better of the two options because it is JavaScript code designed specifically for Flash .swf purposes, and it has been abused to a lesser extent than the **NoScript** tag option.

SWFObject is Flash detection code written in JavaScript that checks whether a browser has the Flash plug-in. If the browser does have the Flash plug-in, the .swf file is displayed secondary to that detection. If the browser does not have the Flash plug-in or the JavaScript to detect it, the primary, alternative content contained within <div> files is displayed instead. The key here is that search engine spiders do not render the JavaScript. They read the primary content in the <div> tags.

The opportunity for abuse is obvious upon viewing the code. This small piece of code is placed within the <head> tags:

```
<script type="text/javascript" src="swfobject.js"></script>
```

In the body of the text, the code looks something like [Figure 6-44](#).

```
<script type="text/javascript" src="swfobject.js"></script>

<div id="flashcontent">
    Text, links, and graphics placed here are replaced by the Flash movie. Search
    engine spiders will read this information, but the browser with an active Flash
    plugin will show the Flash movie instead.
</div>

<script type="text/javascript">
    var so = new SWFObject("whatever.swf", "themovie" , "200", "100", "7", "#336699");
    so.write("flashcontent");
</script>
```

FIGURE 6-44. Information between the <div> HTML tags read by search engine spiders

Search engine spiders will read text, links, and even alt attributes within the <div> tags, but the browser will not display them unless the Flash plug-in isn't installed (about 95% of browsers now have the plug-in) or JavaScript isn't available.

Once again, the key to successfully implementing **SWFObject** is to use it to the letter of the law; leverage it to mirror the content of your Flash .swf file *exactly*. Do not use it to add content, keywords, graphics, or links that are not contained in the file. Remember, a human being will be making the call as to whether your use of **SWFObject** is proper and in accordance with that search engine's guidelines. If you design the outcome to provide the best possible user experience, and your intent is *not* to game the search engines, you are probably OK.

You can download the **SWFObject** JavaScript free of charge at <http://code.google.com/p/swfobject/>. Included in this download is the *flashobject.js* file, which is placed in the same directory as the web pages upon which the corresponding calling code resides.

NoScript. The **NoScript** tag has been abused in “black hat” SEO attempts so frequently that caution should be taken when using it. Just as **SWFObject** and **DIV** tags can be misused for link

and keyword stuffing, so too can the `NoScript` tag. Certain companies have promoted the misuse of the `NoScript` tag widely; consequently there have been many more problems with their use.

With that being said, conservative and proper use of `NoScript` tags specifically with Flash `.swf` files can be an acceptable and good way to get content mirrored to a Flash file read by search engine spiders. As it is with `SWFObject` and corresponding `DIV` tags, content must echo that of the Flash `.swf` movie exactly. Do not use them to add content, keywords, graphics, or links that are not in the movie. Again, it is a human call as to whether a site or individual page is banned for the use or misuse of `NoScript` tags.

You use `NoScript` tags with Flash `.swf` files in the following manner:

```
<script type="text/javascript" src="YourFlashFile.swf"></script>
```

Followed at some point afterward by:

```
<noscript>
<H1>Mirror content in Flash file here.</H1>
<p>Any content within the NoScript tags will be read by the search engine
spiders, including links
http://www.mirroredlink.com, graphics, and corresponding Alt attributes.
</noscript>
```

For browsers that do not have JavaScript installed or functioning, content alternatives to JavaScript-required entities are displayed. So, for use with Flash `.swf` files, if a browser does not have JavaScript and therefore cannot display Flash, it displays instead the content inside the `NoScript` tags. This is a legitimate, graceful degradation design. For SEO purposes, as it is with `SWFObject`, the search engine spiders do not render the JavaScript and do read the content contained in the HTML. Here, it is the content in the `NoScript` tags.

Scalable Inman Flash Replacement (sIFR)

sIFR is a technique that uses JavaScript to read in HTML text and render it in Flash instead. The essential fact to focus on here is that the method guarantees that the HTML content and the Flash content are identical. One great use for this is to render headline text in an anti-aliased font (this is the purpose for which sIFR was designed). This can provide a great improvement in the presentation of your site.

At a Search Engine Marketing New England (SEMNE) event in July 2007, Dan Crow, head of Google's Crawl team, said that as long as this technique is used in moderation, it is OK.

However, extensive use of sIFR could be interpreted as a poor site quality signal. Since sIFR was not really designed for large-scale use, such extensive use would not be wise in any event.

JavaScript and AJAX

As we mentioned previously, the search engines cannot see any content that does not appear immediately in the source code for the web page. Content and links that are dynamically retrieved in response to a user action or some other event and then embedded in the page

cannot be seen by the search engines. However, content that is embedded in the HTML of the page, and JavaScript that is simply used to render it, can be seen by the search engines. There is also some evidence that Google follows links implemented in JavaScript, as detailed in the article on SEOMoz at <http://www.seomoz.org/ugc/new-reality-google-follows-links-in-javascript-4930>.

AJAX implementations can also be very challenging. The main challenge is that most implementations render new content on the page, but the URL does not change. To build AJAX applications that present indexable content to the search engines, make sure to create new URLs that point to the various versions of the content, and then make sure you link to those locations from somewhere on your site.

Use of the `NoScript` tags to render alternative content for the benefit of the search engines is useful in addressing basic JavaScript applications, but less useful with AJAX, since there may be an infinite (or very large) number of different types of output from the tool, and it would not be feasible to render them all within the `NoScript` tags.

Best Practices for Multilanguage/Country Targeting

Many businesses target multiple countries with their websites and need answers to questions such as: do you put the information for your products or services all on the same domain? Do you obtain multiple domains? Where do you host the site(s)? It turns out that there are SEO factors, as well as basic marketing questions, that affect the answers. There are also non-SEO factors, such as the tax implications of what you do; for some TLDs you can get them only by having a local physical presence (e.g., France requires this to get a .fr domain).

Targeting a Specific Country

Starting with the basics of international targeting, it is important to let the search engines know where your business is based in as many ways as possible. These might include:

- A country-specific TLD (ccTLD) for your domain (e.g., .co.uk)
- Hosting in the local country
- Physical local address in plain text on every page of your site
- Google Webmaster Central geotargeting setting
- Verified address with Google Maps
- Links from in-country websites
- Use of the local language on the website

If you are starting from scratch, getting these all lined up will give you the best possible chance of ranking in the local country you are targeting.

Problems with Using Your Existing Domain

You may ask why you cannot leverage your domain weight to target the new territory rather than starting from scratch—in other words, create multiple versions of your site and determine where the user is in the world before either delivering the appropriate content or redirecting him to the appropriate place in the site (or even to a subdomain hosted in the target country).

The problem with this approach is that the search engines spider from the United States, their IP addresses will be in the United States in your lookup, and they will therefore be delivered in U.S. content. This problem is exacerbated if you are going even further and geodelivering different language content, as only your English language content will be spidered unless you cloak for the search engine bots.

This kind of IP delivery is therefore a bad idea. You should make sure you do not blindly geodeliver content based on IP address as you will ignore many of your markets in the search engines' eyes.

The Two Major Approaches

The best practice remains one of two approaches, depending on the size and scale of your operations in the new countries and how powerful and established your .com domain is.

If you have strong local teams and/or (relatively speaking) less power in your main domain, launching independent local websites geotargeted as described earlier (hosted locally, etc.) is a smart move in the long run.

If, on the other hand, you have only centralized marketing and PR and/or a strong main domain, you may want to create localized versions of your content either on subdomains (.uk, .au, etc.) or in subfolders (/uk/, /au/, etc.), with the preference being for the use of subdomains so that you can set up local hosting.

Both the subdomains and the subfolders approaches allow you to set your geotargeting option in Google Webmaster Central, and with either method you have to be equally careful of duplicate content across regions. In the subdomain example, you can host the subdomain locally, while in the subfolder case, more of the power of the domain filters down.

Unfortunately, the Webmaster Tools' geotargeting option doesn't work nearly as well as you'd hope to geotarget subfolders. The engines will consider hosting and ccTLDs, along with the geographic location of your external link sources, to be stronger signals than the manual country targeting in the tools. In addition, people in other countries (e.g., France) don't like to click on ".com" or ".org" TLDs—they prefer ".fr". This extends to branding and conversion rates too—web users in France like to buy from websites in France that end in ".fr".

Multiple-Language Issues

An entire treatise could be written on handling multilanguage content as the search engines themselves are rapidly evolving in this field, and tactics are likely to change dramatically in the near future. Therefore, this section will focus on providing you with the fundamental components of successful multilanguage content management.

Here are best practices for targeting the search engines as of this writing, using Spanish and English content examples:

- Content in Spanish and English serving the same country:
 - Create a single website with language options that change the URL by folder structure; for example, *http://www.yourdomain.com* versus *http://www.yourdomain.com/esp/*.
 - Build links from Spanish and English language sites to the respective content areas on the site.
 - Host the site in the country being served.
 - Register the appropriate country domain name (for the United States, .com, .net, and .org are appropriate, whereas in Canada using .ca or in the United Kingdom using .co.uk is preferable).
- Content in Spanish and English targeting multiple countries:
 - Create two separate websites, one in English targeting the United States (or the relevant country) and one in Spanish targeting a relevant Spanish-speaking country.
 - Host one site in the United States (for English) and the other in the relevant country for the Spanish version.
 - Register different domains, one using U.S.-targeted domain extensions and one using the Spanish-speaking country's extension.
 - Acquire links from the United States to the English site and links from the Spanish-speaking country to that site.
- Content in Spanish targeting multiple countries:
 - Create multiple websites (as mentioned earlier) targeting each specific country.
 - Register domains using the appropriate country TLD and host in each country separately.
 - Rewrite the content of each site to be unique, as duplicate content problems will arise despite the separation of targeting.
 - When possible, have native speakers fluent in the specific region's dialect write the site content for each specific country.
 - Obtain in-country links to your domains.

Although some of these seem counterintuitive, the joint issues of search engines preferring to show content hosted in and on a country-specific domain name combined with duplicate

content problems make for these seemingly illogical suggestions. Creating multiple websites is never ideal due to the splitting of link equity, but in the case of international targeting, it is often hard or even impossible to rank a U.S.-hosted .com address for foreign language content in another country.

Conclusion

A search-engine-friendly website is key to SEO success. In the next chapter (on link building), we will demonstrate that links are also a critical piece of the SEO puzzle—particularly when targeting highly competitive terms. However, if you have not made your site crawler-friendly and optimized, link building will essentially be a wasted effort. A website built, from the ground up, to optimal specifications for crawler accessibility and top rankings is the foundation from which you will build all SEO initiatives. From this solid foundation, even the loftiest of goals are within reach.

CHAPTER SEVEN

Creating Link-Worthy Content and Link Marketing

LINKS ARE THE MAIN DETERMINANTS OF RANKING BEHAVIOR. Both site architecture and content are big players in achieving search engine friendliness, but when it comes to the mechanism for ordering results, external links are the critical metric.

Links are also multidimensional in their impact. For example, the strength of a site's inbound links determines how frequently and how deeply a site is crawled.

In addition, each link's *context* (its location within the page, anchor text, surrounding text, etc.) is taken into account when determining relevance. And links can convey trust, helping to overcome potential spam designations (which can torpedo a site's SEO success).

There are two critical points to remember regarding link building:

- Link building is a fundamental part of SEO. Unless you have an enormously powerful brand (one that attracts links without effort), you will fail without it.
- Link building should never stop. It is an ongoing part of marketing your website.

How Links Influence Search Engine Rankings

The concept of using links as a way to measure a site's importance was first made popular by Google with the implementation of its PageRank algorithm (others had previously written

about it, but Google's rapidly increasing user base popularized it). In simple terms, each link to a web page is a vote for that page, and the page with the most votes wins.

The key to this concept is the notion that links represent an “editorial endorsement” of a web document. Search engines rely heavily on editorial votes. However, as publishers learned about the power of links, some publishers started to manipulate links through a variety of methods. This created situations in which the intent of the link was not editorial in nature, and led to many algorithm enhancements, which we will discuss in this chapter.

To help you understand the origins of link algorithms, the underlying logic of which is still in force today, let's take a look at the original PageRank algorithm in detail.

The Original PageRank Algorithm

The PageRank algorithm was built on the basis of the original [PageRank thesis](#) authored by Sergey Brin and Larry Page while they were undergraduates at Stanford University.

In the simplest terms, the paper states that each link to a web page is a vote for that page. However, votes do not have equal weight. So that you can better understand how this works, we'll explain the PageRank algorithm at a high level. First, all pages are given an innate but tiny amount of PageRank, as shown in [Figure 7-1](#).

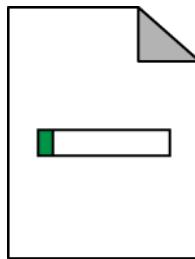


FIGURE 7-1. Some PageRank for every page

Pages can then increase their PageRank by receiving links from other pages, as shown in [Figure 7-2](#).

How much PageRank can a page pass on to other pages through links? That ends up being less than the page's PageRank. In [Figure 7-3](#) this is represented by $f(x)$, meaning that the passable PageRank is a function of x , the total PageRank.

If this page links to only one other page, it passes all of its PageRank to that page, as shown in [Figure 7-4](#), where Page B receives all of the passable PageRank of Page A.

However, the scenario gets more complicated because pages will link to more than one other page. When that happens the passable PageRank gets divided among all the pages receiving

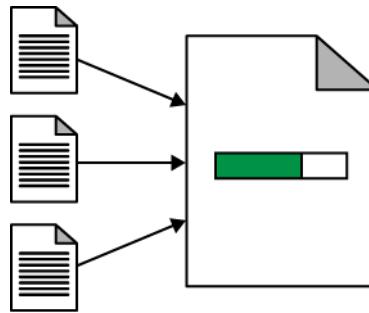


FIGURE 7-2. Pages receiving more PageRank through links

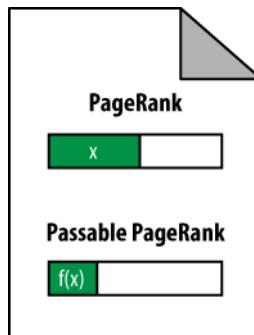


FIGURE 7-3. Some of a page's PageRank passable to other pages

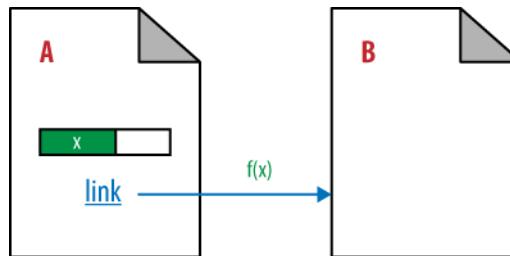


FIGURE 7-4. Passing of PageRank through a link

links. We show that in Figure 7-5, where Page B and Page C each receive half of the passable PageRank of Page A.

In the original PageRank formula, link weight is divided equally among the number of links on a page. This undoubtedly does not hold true today, but it is still valuable in understanding

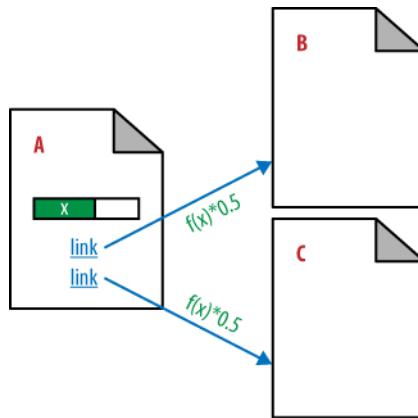


FIGURE 7-5. Simple illustration of how PageRank is passed

the original intent. Now take a look at [Figure 7-6](#), which depicts a more complex example that shows PageRank flowing back and forth between pages that link to one another.

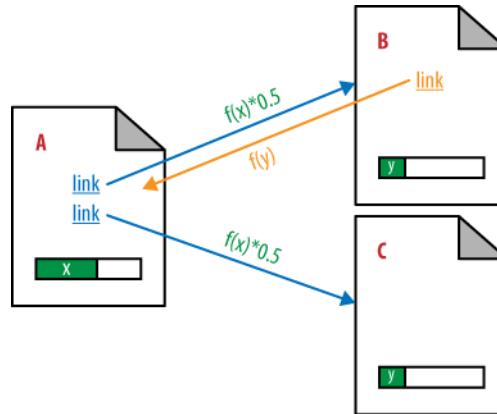


FIGURE 7-6. Cross-linking between pages

Cross-linking makes the PageRank calculation much more complex. In [Figure 7-6](#), Page B now links back to Page A and passes some PageRank, $f(y)$, back to Page A. [Figure 7-7](#) should give you a better understanding of how this affects the PageRank of all the pages.

The key observation here is that when Page B links to Page A to make the link reciprocal, the PageRank of Page A (x) becomes dependent on $f(y)$, the passable PageRank of Page B, which happens to be dependent on $f(x)$!. In addition, the PageRank that Page A passes to Page C is also impacted by the link from Page B to Page A. This makes for a very complicated situation

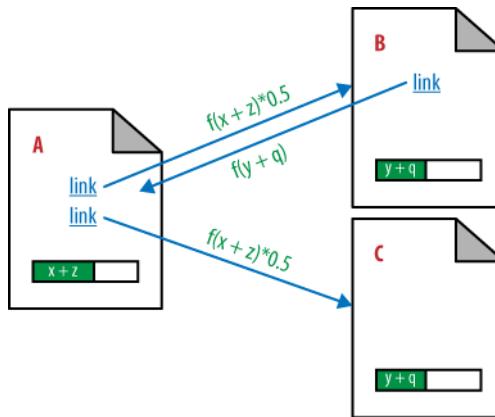


FIGURE 7-7. Iterative PageRank calculations

where the calculation of the PageRank of each page on the Web must be determined by recursive analysis.

We have defined new parameters to represent this: q , which is the PageRank that accrues to Page B from the link that it has from Page A (after all the iterative calculations are complete); and z , which is the PageRank that accrues to Page A from the link that it has from Page B (again, after all iterations are complete).

The scenario in [Figure 7-8](#) adds additional complexity by introducing a link from Page B to Page D. In this example, pages A, B, and C are internal links on one domain, and Page D represents a different site (shown as Wikipedia). In the original PageRank formula, internal and external links passed PageRank in exactly the same way. This became exposed as a flaw because publishers started to realize that links to other sites were “leaking” PageRank away from their own site, as you can see in [Figure 7-8](#).

Because Page B links to Wikipedia, some of the passable PageRank is sent there, instead of to the other pages that Page B is linking to (Page A in our example). In [Figure 7-8](#), we represent that with the parameter w , which is the PageRank not sent to Page A because of the link to Page D.

The PageRank “leak” concept presented a fundamental flaw in the algorithm once it became public. Like Pandora’s Box, once those who were creating pages to rank at Google investigated PageRank’s founding principles, they would realize that linking out from their own sites would cause more harm than good. If a great number of websites adopted this philosophy, it could negatively impact the “links as votes” concept and actually damage Google’s potential. Needless to say, Google corrected this flaw to its algorithm. As a result of these changes, worrying about PageRank leaks is not recommended. Quality sites should link to other relevant quality pages around the Web.

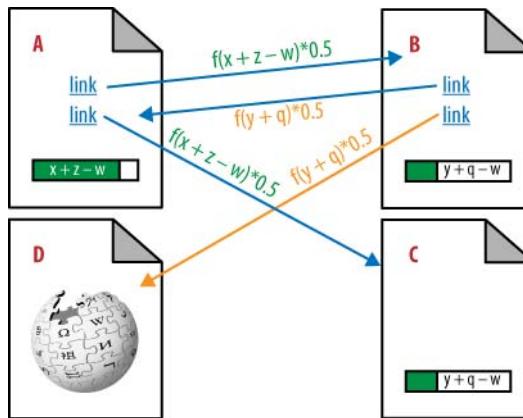


FIGURE 7-8. PageRank being leaked

Even after these changes, internal links from pages still pass some PageRank, so they still have value, as shown in Figure 7-9.

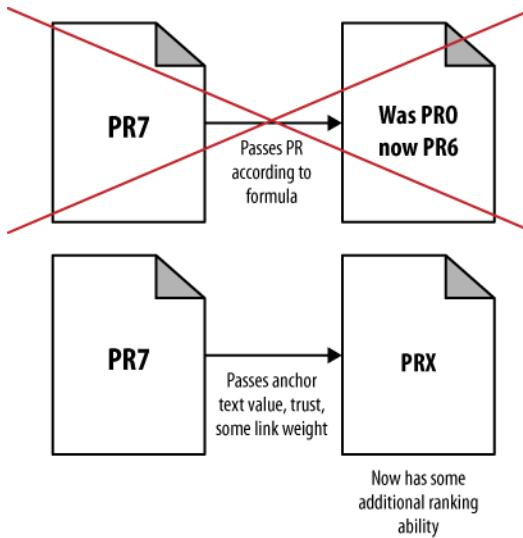


FIGURE 7-9. Internal links still passing some PageRank

Google has changed and refined the PageRank algorithm many times. However, familiarity and comfort with the original algorithm is certainly beneficial to those who practice optimization of Google results.

Additional Factors That Influence Link Value

Classic PageRank isn't the only factor that influences the value of a link. In the following subsections, we discuss some additional factors that influence the value a link passes.

Anchor text

Anchor text refers to the clickable part of a link from one web page to another. As an example, [Figure 7-10](#) shows a snapshot of a part of the Alchemist Media Home Page at <http://www.alchemistmedia.com>.

[Search Engine Optimization \(SEO\) & Web Site Promotion Link #1](#)

SEO is not an easy discipline to understand, let alone master. There are many suggested shortcuts that search engine consultants and optimization firms recommend. However, in the long run, these approaches can be counterproductive. If your approach is too narrow in focus, and at worst, some of the optimization techniques used can also result in your site's pages being penalized and your rankings lowered, or in your site being removed from a search engine's index entirely. Learn to take a cautious yet comprehensive approach to SEO by engaging only in tactics that result in a better, more relevant web site for both end users and search engines. We move the focus away from one-shot, quick-fix optimization towards helping you develop an overall Organic Search Methodology.

[Pay Per Click Search Engine Marketing Link #2](#)

PPC advertising allows you to pay on a cost-per click basis to have your text ads listed in the search results pages of the major search engines. These listings are keyword-based, meaning that you can control which keyword search results pages you would like your ad to appear on. Pay per click listings for each keyword are ranked on a combination of the bid (amount you are willing to pay for each click) and the "Quality Score" of your site as it relates to each keyword in your campaign. From keyword research to landing page conversion, there are many factors that can help maximize your return on investment.

[SEO Web Site Design Link #3](#)

One of the most important factors when attempting to rank your site in the search engines is designing your web site in such a way that both the search engines and your web site visitors are able to quickly, easily, and thoroughly access your site content from the landing page onward. Successful SEO web site design requires a fundamental knowledge and understanding of how search engines "see" your web page content, and how they determine its relevance. We show you how to think like a search engine as well as a searcher.

FIGURE 7-10. Anchor text: a strong ranking element

The anchor text for Link #3 in [Figure 7-10](#) is "SEO Web Site Design". The search engine uses this anchor text to help it understand what the page receiving the link is about. As a result, the search engine will interpret Link #3 as saying that the page receiving the link is about "SEO Web Site Design".

The impact of anchor text can be quite powerful. For example, if you link to a web page that has no search-engine-visible content (perhaps it is an all-Flash site), the search engine will still look for signals to determine what the page is about. Inbound anchor text becomes the primary driver in determining the relevance of a page in that scenario.

The power of anchor text also resulted in SEOs engaging in *Google Bombing*. The idea is that if you link to a given web page from many places with the same anchor text, you can get that page to rank for queries related to that anchor text, even if the page is unrelated.

One notorious Google Bomb was a campaign that targeted the Whitehouse.gov biography page for George W. Bush with the anchor text *miserable failure*. As a result, that page ranked #1 for searches on *miserable failure* until Google tweaked its algorithm to reduce the effectiveness of this practice.

However, this still continues to work in Yahoo! Search (as of May 2009), as shown in [Figure 7-11](#).



FIGURE 7-11. Anchor text making pages rank for unrelated terms

President Obama has crept in here too, largely because of a redirect put in place by the White House's web development team.

Relevance

Links that originate from sites/pages on the same topic as the publisher's site, or on a closely related topic, are worth more than links that come from a site on an unrelated topic.

Think of the relevance of each link being evaluated in the specific context of the search query a user has just entered. So, if the user enters *used cars in Phoenix* and the publisher has a link to the Phoenix used cars page that is from the Phoenix Chamber of Commerce, that link will reinforce the search engine's belief that the page really does relate to Phoenix.

Similarly, if a publisher has another link from a magazine site that has done a review of used car websites, this will reinforce the notion that the site should be considered a used car site. Taken in combination, these two links could be powerful in helping the publisher rank for "used cars in Phoenix".

Authority

This has been the subject of much research. One of the more famous papers, written by Apostolos Gerasoulis and others at Rutgers University and titled "DiscoWeb: Applying Link Analysis to Web Search" (<http://www.cse.lehigh.edu/~brian/pubs/1999/www8/>), became the basis of the Teoma algorithm, which was later acquired by AskJeeves and became part of the Ask algorithm.

What made this unique was the focus on evaluating links on the basis of their relevance to the linked page. Google's original PageRank algorithm did not incorporate the notion of topical relevance, and although Google's algorithm clearly does do this today, Teoma was in fact the first to offer a commercial implementation of link relevance.

Teoma introduced the notion of *hubs*, which are sites that link to most of the important sites relevant to a particular topic, and *authorities*, which are sites that are linked to by most of the sites relevant to a particular topic.

The key concept here is that each topic area that a user can search on will have authority sites specific to that topic area. The authority sites for used cars are different from the authority sites for baseball.

Refer to [Figure 7-12](#) to get a sense of the difference between hub and authority sites.

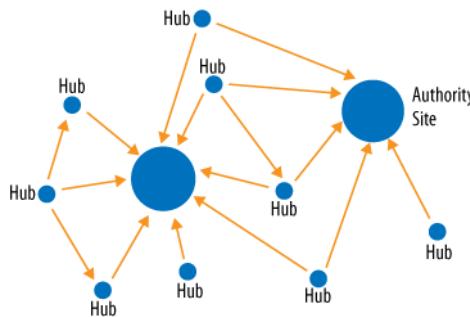


FIGURE 7-12. Hubs and authorities

So, if the publisher has a site about used cars, it seeks links from websites that the search engines consider to be authorities on used cars (or perhaps more broadly, on cars). However, the search engines will not tell you which sites they consider authoritative—making the publisher's job that much more difficult.

The model of organizing the Web into topical communities and pinpointing the hubs and authorities is an important model to understand (read more about it in Mike Grehan's paper, "Filthy Linking Rich!" at http://www.search-engine-book.co.uk/filthy_linking_rich.pdf). The best link builders understand this model and leverage it to their benefit.

Trust

Trust is distinct from authority. Authority, on its own, doesn't sufficiently take into account whether the linking page or the domain is easy or difficult for spammers to infiltrate. Trust, on the other hand, does.

Evaluating the trust of a website likely involves reviewing its link neighborhood to see what other trusted sites link to it. More links from other trusted sites would convey more trust.

In 2004, Yahoo! and Stanford University published a paper titled “Combating Web Spam with TrustRank” (<http://www.vldb.org/conf/2004/RS15P3.PDF>). The paper proposed starting with a trusted seed set of pages (selected by manual human review) to perform PageRank analysis, instead of a random set of pages as was called for in the original PageRank thesis.

Using this tactic removes the inherent risk in using a purely algorithmic approach to determining the trust of a site, and potentially coming up with false positives/negatives.

The trust level of a site would be based on how many clicks away it is from seed sites. A site that is one click away accrues a lot of trust; two clicks away, a bit less; three clicks away, even less; and so forth. [Figure 7-13](#) illustrates the concept of TrustRank.

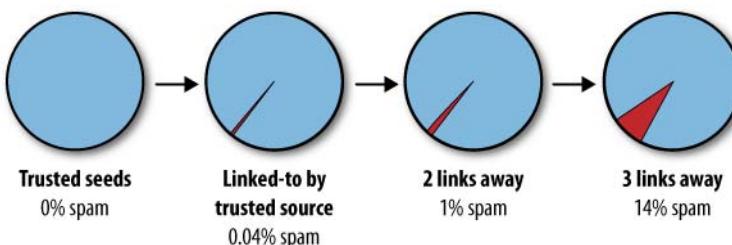


FIGURE 7-13. TrustRank illustrated

The researchers of the TrustRank paper also authored a paper describing the concept of *spam mass* (<http://ilpubs.stanford.edu:8090/697/1/2005-33.pdf>). This paper focuses on evaluating the effect of spammy links on a site’s (unadjusted) rankings. The greater the impact of those links, the more likely the site itself is spam. A large percentage of a site’s links being purchased is seen as a spam indicator as well. You can also consider the notion of *reverse TrustRank*, where linking to spammy sites will lower a site’s TrustRank.

It is likely that Google, Yahoo!, and Bing all use some form of trust measurement to evaluate websites, and that this trust metric can be a significant factor in rankings. For SEO practitioners, getting measurements of trust can be difficult. Currently, mozTrust from SEOMoz’s Linkscape is the only publicly available measured estimation of a page’s TrustRank.

How Search Engines Use Links

The search engines use links primarily to discover web pages, and to count the links as votes for those web pages. But how do they use this information once they acquire it? Let’s take a look:

Index inclusion

Search engines need to decide what pages to include in their index. Discovering pages by crawling the Web (following links) is one way they discover web pages (the other is through the use of XML Sitemap files). In addition, the search engines do not include pages

that they deem to be of low value because cluttering their index with those pages will not lead to a good experience for their users. The cumulative link value, or link juice, of a page is a factor in making that decision.

Crawl rate/frequency

Search engine spiders go out and crawl a portion of the Web every day. This is no small task, and it starts with deciding where to begin and where to go. Google has publicly indicated that it starts its crawl in PageRank order. In other words, it crawls PageRank 10 sites first, PageRank 9 sites next, and so on. Higher PageRank sites also get crawled more deeply than other sites. It is likely that other search engines start their crawl with the most important sites first as well.

This would make sense, because changes on the most important sites are the ones the search engines want to discover first. In addition, if a very important site links to a new resource for the first time, the search engines tend to place a lot of trust in that link and want to factor the new link (vote) into their algorithms quickly.

Ranking

Links play a critical role in ranking. For example, consider two sites where the on-page content is equally relevant to a given topic. Perhaps they are the shopping sites Amazon.com and (the less popular) [JoesShoppingSite.com](#).

The search engine needs a way to decide who comes out on top: Amazon or Joe. This is where links come in. Links cast the deciding vote. If more sites, and more important sites, link to it, it must be more important, so Amazon wins.

Further Refining How Search Engines Judge Links

Many aspects are involved when you evaluate a link. As we just outlined, the most commonly understood ones are authority, relevance, trust, and the role of anchor text. However, other factors also come into play.

Additional Link Evaluation Criteria

In the following subsections, we discuss some of the more important factors search engines consider when evaluating a link's value.

Source independence

A link from your own site back to your own site is, of course, not an independent editorial vote for your site. Put another way, the search engines assume that you will vouch for your own site.

Think about your site as having an accumulated total link juice based on all the links it has received from third-party websites, and your internal linking structure as the way you allocate

that juice to pages on your site. Your internal linking structure is incredibly important, but it does little if anything to build the total link juice of your site.

In contrast, links from a truly independent source carry much more weight. Extending this notion a bit, it may be that you have multiple websites. Perhaps they have common data in the Whois records (such as the IP address or contact information). Search engines can use this type of signal to treat cross-links between those sites that are more like internal links than inbound links earned by merit.

Even if you have different Whois records for the websites but they all cross-link to each other, the search engines can detect this pattern easily. Keep in mind that a website with no independent third-party links into it has no link power to vote for other sites.

If the search engine sees a cluster of sites that heavily cross-link and many of the sites in the cluster have no or few incoming links to them, the links from those sites may well be ignored.

Conceptually, you can think of such a cluster of sites as a single site. Cross-linking to them can be algorithmically treated as a single site, with links between them not adding to the total link juice score for each other. The cluster would be evaluated based on the inbound links to the cluster.

Of course, there are many different ways such things could be implemented, but one thing that would not have SEO value is to build a large number of sites, just to cross-link them with each other.

Linking domains

Getting a link editorially given to your site from a third-party website is always a good thing. But if more links are better, why not get links from every page of these sites if you can? In theory, this is a good idea, but search engines do not count multiple links from a domain cumulatively.

In other words, 100 links from one domain are not as good as one link from 100 domains, if you assume that all other factors are equal. The basic reason for this is that the multiple links on one site most likely represent one editorial vote. In other words, it is highly likely that one person made the decision. Furthermore, a sitewide link is more likely to have been paid for.

So, multiple links from one domain are still useful to you, but the value per added link in evaluating the importance of your website diminishes as the quantity of links goes up. One hundred links from one domain might carry the total weight of one link from 10 domains. One thousand links from one domain might not add any additional weight at all.

Our experimentation suggests that anchor text is not treated quite the same way, particularly if the links are going to different pages on your site. In other words, those multiple links may communicate link value in a way that is closer to linear. More links might not mean more importance, but the editorial vote regarding the topic of a page through the anchor text remains interesting data for the search engines, even as the link quantity increases.

Think about the number of unique linking domains as a metric for your site, or for any site you are evaluating. If you have a choice between getting a new link from a site that already links to you as opposed to getting a new link from a domain that currently does not link to you, go with the latter choice nearly every time (we will discuss this process in more detail later in this chapter).

Source diversity

Getting links from a range of sources is also a significant factor. We already discussed two aspects of this: getting links from domains you do not own, and getting links from many different domains. However, there are many other aspects of this.

For example, perhaps all your links come from blogs that cover your space. This ends up being a bit unbalanced. You can easily think of other types of places where you could get links: directories, social media sites, university sites, media websites, social bookmarking sites, and so on.

You can think about implementing link-building campaigns in many of these different sectors as *diversification*. There are several good reasons for doing this.

One reason is that the search engines value this type of diversification. If all your links come from a single class of sites, the reason is more likely to be manipulation, and search engines do not like that. If you have links coming in from multiple types of sources, that looks more like you have something of value.

Another reason is that search engines are constantly tuning and tweaking their algorithms. If you had all your links from blogs and the search engines made a change that significantly reduced the value of blog links, that could really hurt your rankings. You would essentially be hostage to that one strategy, and that's not a good idea either.

Temporal factors

Search engines also keep detailed data on when they discover the existence of a new link, or the disappearance of a link. They can perform quite a bit of interesting analysis with this type of data. Here are some examples:

When did the link first appear?

This is particularly interesting when considered in relationship to the appearance of other links. Did it happen immediately after you received that link from the *New York Times*?

When did the link disappear?

Some of this is routine, such as links that appear in blog posts that start on the home page of a blog and then get relegated to archive pages over time. However, perhaps it is after you rolled out a new major section on your site, which could be an entirely different type of signal.

How long has the link existed?

You can potentially count a link for less if it has been around for a long time. Whether you choose to count it for more or less could depend on the authority/trust of the site providing the link, or other factors.

How quickly were the links added?

Did you go from one link per week to 100 per day, or vice versa? Such drastic changes in the rate of link acquisition could also be a significant signal. Whether it is a bad signal or not depends. For example, if your site is featured in major news coverage it could be good. If you start buying links by the thousands it could be bad. Part of the challenge for the search engines is to determine how to interpret the signal.

Context/relevance

Although anchor text is a major signal regarding the relevance of a web page, search engines look at a much deeper context than that. They can look at other signals of relevance. Here are some examples of those:

Nearby links

Do the closest links on the page point to closely related, high-quality sites? That would be a positive signal to the engines, as your site could be seen as high-quality by association. Alternatively, if the two links before yours are for Viagra and a casino site, and the link after yours points to a porn site, that's not a good signal.

Page placement

Is your link in the main body of the content? Or is it off in a block of links at the bottom of the right rail of the web page? Better page placement can be a ranking factor. This is also referred to as *prominence*, which has application in on-page keyword location as well.

Nearby text

Does the text immediately preceding and following your link seem related to the anchor text of the link and the content of the page on your site that it links to? If so, that could be an additional positive signal. This is also referred to as *proximity*.

Closest section header

Search engines can also look more deeply at the context of the section of the page where your link resides. This can be the nearest header tag, or the nearest text highlighted in bold, particularly if it is implemented like a header (two to four boldface words in a paragraph by themselves).

Overall page context

The relevance and context of the linking page are also factors in rankings. If your anchor text, surrounding text, and the nearest header are all related, that's good. But if the overall context of the linking page is also closely related, that's better still.

Overall site context

Last is the notion of the context of the entire site that links to you (or perhaps even just the section of the site that links to you). For example, if hundreds of pages are relevant to your topic and you receive your link from a relevant page, with relevant headers, nearby text, and anchor text, these all add to the impact, more than if there happens to be only one relevant page on the site.

Source TLDs

Indications are that there is no preferential treatment for certain top-level domains (TLDs), such as .edu, .gov, and .mil. It is a popular myth that these TLDs are a positive ranking signal, but it does not make sense for search engines to look at it so simply.

Matt Cutts, the head of the Google webspam team, commented on this in an interview with Stephan Spencer (<http://www.stephanspencer.com/search-engines/matt-cutts-interview>):

There is nothing in the algorithm itself, though, that says: oh, .edu—give that link more weight.

And:

You can have a useless .edu link just like you can have a great .com link.

There are many forums, blogs, and other pages on .edu domains that spammers easily manipulate to gain links to their sites. For this reason, search engines cannot simply imbue a special level of trust or authority to a site because it is an .edu domain. To prove this to yourself, simply search for *buy viagra site:edu* to see how spammers have infiltrated .edu pages.

However, it is true that .edu domains are often authoritative. But this is a result of the link analysis that defines a given college or university as a highly trusted site on one or more topics. The result is that there can be (and there are) domains that are authoritative on one or more topics on some sections of their site, and yet can have another section of their site that spammers are actively abusing.

Search engines deal with this problem by varying their assessment of a domain's authority across the domain. The publisher's *http://yourdomain.com/usecdars* section may be considered authoritative on the topic of used cars, but *http://yourdomain.com/newcars* might not be authoritative on the topic of new cars.

One technique that *link brokers* (companies that sell links) use is the notion of *presell pages*. These are pages on an authoritative domain for which the link broker has obtained the right to place and sell ad copy and links to advertisers. The link broker pays the domain owner a sum of money, or a percentage of the resulting revenue, to get control over these pages.

For example, the link broker may negotiate a deal with a major university enabling it to place one or more pages on the university's website. The links from this page do have some of the inherent value that resides in the domain. However, the presell pages probably don't have many (if any) links from other pages on the university site or from other websites. As with

other forms of purchasing links, presell pages are considered spam by search engines, and pursuing these types of links is a high-risk tactic.

Ultimately, every site and every page on every site gets evaluated for the links they have, on a topic-by-topic basis. Further, each section of a site and each page also get evaluated on this basis. A certain link profile gives a page more authority on a given topic, making that page likely to rank higher on queries for that topic, and providing that page with more valuable links that it could then give to other websites related to that topic.

Link and document analysis combine and overlap, resulting in hundreds of factors that can be individually measured and filtered through the search engine algorithms (the set of instructions that tells the engines what importance to assign to each factor). The algorithms then determine scoring for the documents and (ideally) list results in decreasing order of relevance and importance (rankings).

Determining a Link's Value

Putting together a link campaign generally starts with researching sites that would potentially link to the publisher's site and then determining the relative value of each potential linker. Although there are many metrics for evaluating a link, as we just discussed, as an individual link builder many of those data items are hard to determine (e.g., when a link was first added to a site).

It is worth taking a moment to outline an approach that you can use today, with not too much in the way of specialized tools. Here are factors you can look at:

- The PageRank of the home page of the site providing the link. Note that Google does not publish a site's PageRank, just the PageRank for individual pages. It is common among SEO practitioners to use the home page of a site as a proxy for the site's overall PageRank, since a site's home page typically garners the most links. You can also use the Domain mozRank from SEOMoz's Linkscape tool to get a third-party approximation of domain PageRank.
- The perceived authority of the site. Although there is a relationship between authority and PageRank, they do not have a 1:1 relationship. Authority relates to how the sites in a given market space are linked to by other significant sites in the same market space, whereas PageRank measures aggregate raw link value without regard to the market space. So, higher-authority sites will *tend* to have higher PageRank, but this is not absolutely the case.
- The PageRank of the linking page.
- The perceived authority of the linking page.
- The number of outbound links on the linking page. This is important because the linking page can vote its passable PageRank for the pages to which it links, but each page it links

to consumes a portion of that PageRank, leaving less to be passed on to other pages. This can be expressed mathematically as follows:

For a page with passable PageRank n and with r outbound links:

$$\text{Passed PageRank} = n/r$$

This is a rough formula, but the bottom line is that the more outbound links a page has, the less valuable a link from that page will be.

- The relevance of the linking page and the site.

Organizing this data in a spreadsheet, or at least being consciously aware of these factors when putting together a link-building campaign, is a must. For many businesses, there will be many thousands of prospects in a link campaign. With a little forethought you can prioritize these campaigns to bring faster results.

The Psychology of Linking

Since we have already established the importance of links, the next logical question is how to go about getting them.

Why Are Links Created?

It is important to step back and examine why links are created in the first place. Why did that person decide to link to that particular website? There are many possible reasons:

1. She was paid for the link. Although this is a perfectly legitimate reason, in the search engine's eyes it carries no editorial value (and search engines may even penalize sites for linking or attracting links in this fashion).
2. Links were traded between sites. Also called *reciprocal linking*, the practice of trading links between sites is popular. However, search engines view this as barter and therefore as having limited editorial value.
3. Something on your site triggered an emotional reaction from the publisher, causing it to link to your site. For example, perhaps your site had the funniest cartoon the publisher ever saw, or it offered an inflammatory political opinion.
4. The publisher may see something of value on your site and wants its site visitors to know about it. The majority of the highest-value links are given for this reason.
5. A business relationship comes into play. For example, you may have a network of distributors and resellers for your product. Do they all link back to you?

How Can Sites Approach Getting Links?

The keys to getting links are points 3, 4, and 5 in the preceding list. Understanding these link triggers is the key to successful link building, as follows:

- Since creating emotional reactions can result in links, building content that plays to the emotions of potential linkers can be a powerful tool for obtaining links. Sometimes this process is referred to as *link baiting*. One key decision for a business is whether such an approach to link building is consistent with the company's brand image, but even a conservative brand can often find worthy link bait topics to be acceptable.
- Create quality reference material. Providing unique and powerful information to users can be a great way to get links, particularly if you can make sure the publishers of authoritative sites in your market space learn about what you have created, including why it would be of value to their website visitors.
- Leverage business relationships. In our example, we suggested that you might be someone who has a network of resellers. If this is your business model, having a link back to you as a standard term in your reseller agreement is sound business.

Of course, this can go much further than this simple concept. There are many situations in which a business relationship or a personal relationship can be a key to getting some additional links.

Types of Link Building

There are many different link-building tactics—too many to list in this book. This section will examine in depth some of the more common ones.

Using Content to Attract Links

In natural link building, the publisher must provide compelling content. Publishers need a good reason to provide links to another site, and it is not something they do frivolously. Superior content or tools are the key to obtaining such links.

Aggressive publishers can even let their content strategy be guided by their link-building strategy. This is not to say that they should let their link-building strategy drive their business strategy.

Normally, however, there are many different types of content a site could produce. The concept is simply to identify the link-building targets and what content will most resonate with the publisher of the target sites, and then tweak the content plan accordingly. Keyword research can also help identify content related to your target market, and can play a role in identifying topics that may help attract links.

Content is at the heart of achieving link-building nirvana—having a site so good that people discover it and link to it without any effort on the publisher's part. This can be done, but it does require that the publisher create content that truly stands out for the topics that its site covers, and that it thinks about link acquisition in every aspect of the publishing process.

The types of content that can attract links vary by market. Here are a few basic rules that a publisher can follow to maximize its results:

- Use content that helps establish your site as a leading expert on its topic matter. When you produce really high-quality stuff, it builds trust with the user community and increases your chances of getting links.
- Minimize the commercial nature of the content pages. As an extreme example, no one is going to link to a page where the only things they see above the fold are AdSense ad units, even if the content below it is truly awesome. Of course, there are less obvious ways to present too many ads, such as too much advertising in the areas around the content or obtrusive overlays and animations.
- Do not put ads in the content itself or link to purely commercial pages unless such pages really merit a link based on the content. No one wants to link to a commercial.
- Do not disguise the relationship between the content and the commercial part of your site. This is the opposite side of the coin. If you are a commercial site and you hide it altogether, you run the risk of being viewed as deceitful.

You can use many types of content to attract links. Article content, compelling images, videos, widgets/tools, or even developing online games can be effective link-building tactics.

When content is published on your site, you have other decisions to make, such as whether the content goes in a special section or whether it is integrated throughout the site.

As an example of how you can make this decision, an e-tail site that publishes a large catalog of products may not want all (or some of) the pages in its catalog laden with a lot of article content. A site such as this might build a separate section with all kinds of tips, tricks, and advice related to the products it sells.

On the other hand, an advertising-supported site might want to integrate the content throughout the main body of the site.

Marketing Content for Link Acquisition

Content can be marketed in many ways. These include:

Content syndication

A publisher may choose to create content for placement on another site. One reason for this would be to provide the content to another site in return for a link to its site. This can be an effective link-building strategy.

Social media

Social media sites such as [Digg](#), [Reddit](#), [StumbleUpon](#), and [Delicious](#) can be useful in marketing content. We will cover this more in “[Social Networking for Links](#)” [on page 321](#).

Spreading content via blogs

Blogging can also be a great tactic for link building. Bloggers are very social and interactive by nature, and they tend to link back and forth quite freely. As with other forms of social media, it's best to do this by being an active contributor to other blogs through commenting and building relationships. We will discuss the concept of blogging for links in more detail in “[Social Networking for Links](#)” on page 321.

Directories

Directories can be a great way to obtain links. A large number of directories are out there, and they may or may not require cash to be paid to obtain a listing. [Table 7-1](#) lists some examples of quality directories.

TABLE 7-1. List of quality directories

Directory name	Category
DMOZ (open directory project)	General
Yahoo! Directory	General
Librarians' Internet Index	General
Inc.com Recommended Start-Up Resources	Business
Nature.com Recommended Links	Science
About.com	General
BBBOnLine	Business
The Vegetarian Resource Group	Health
PhysOrg.com	Science
Open Source Initiative	Computers
Business.com	Business
Best of the Web	General
JoeAnt.com	General
eHub by Emily Chang	General
FreelIndex	General
Environment Directory	Science
Audioholics Buying Guide	Business
Fast Company Talent & Careers Resource Center	General
Yudkin's Recommended Publicity & Marketing Resources	Business
Wheelock College Recommended Websites	General

Directory name	Category
New Zealand Tourism Online	Recreation
Eat Well Guide	People
SciLinks	Science
American Library Association Great Web Sites for Kids	General
Princeton University Outdoor Action Program Guide to Outdoor Resources on the Web	Recreation
Essential Links to Sports Resources	Recreation
Blog Toplist	General
Blog Catalog	General
makezine.com	Media
Online Ethics Center	Business
I Train Online	Computers
The Library of Economics and Liberty	Business
The TalkOrigins Archive	Science
TESL/TEFL/TESOL/ESL/EFL/ESOL Links	Arts and Humanities
Counterpunch	Government
National Institute of Nursing Research	Health
National Marine Fisheries Service Northwest Regional Office	Government
Sexuality Information and Education Council of the US	Health
Bloggeries	General
U.S. Global Change Research Program	Government
American Society for Quality	Business
RealClearPolitics	Media
ReadWriteThink	Arts and Humanities
American Philosophical Association	Arts and Humanities
Art History Resources	Arts and Humanities
Rethinking Schools Online	General
International Reading Association	Arts and Humanities
David Chalmers Philosophy Links	Arts and Humanities

The key to success in link-building to directories is to identify the high-quality ones and stay away from the poor-quality ones. A good-quality indicator is whether the directory exists for users or for webmasters; if it's the latter, stay away from it.

What search engines want from directories

Here are the essential factors the search engines look for:

- The paid fee is made in payment for an editorial review, not for a link.
- Editors may, at their whim, change the location, title, and description of the listing.
- Editors may reject the listing altogether.
- Regardless of the outcome, the directory keeps the money (even if the publisher doesn't get a listing).
- The directory has a track record of rejecting submissions. The inverse of this, which is more measurable, is that the quality of the sites listed in the directory is high.

Ultimately, "anything for a buck" directories do not enforce editorial judgment, and therefore the listings do not convey value to the search engines.

To take a closer look at this, let's examine some of the key statements from Yahoo!'s Directory Submission Terms:

For websites that do not feature adult content or services, the Yahoo! Directory Submit service costs US\$299 (nonrefundable) for each Directory listing that is submitted.

I understand that there is no guarantee my site will be added to the Yahoo! Directory.

I understand that Yahoo! reserves the right to edit my suggestion and category placement; movement or removal of my site will be done at Yahoo!'s sole discretion.

These statements make it pretty clear that Yahoo! will in fact reject your submission if your site is not a quality site, and it will keep your money.

Classifying directories

You can divide directories into three buckets:

Directories that provide sustainable links

These are directories that comply with the policies as outlined earlier. Most likely, these links will continue to pass link juice for the foreseeable future.

Directories that pass link juice that may not be sustainable

These are directories that do not comply with the policies as outlined earlier. The reason such directories exist is that search engines tend to use an "innocent until proven guilty" approach. So, the search engine must proactively make a determination of guilt before a directory's ability to pass link juice is turned off.

Even so, link juice from these types of directories is probably not going to be passed in the long term.

Directories that do not pass link juice

These are the directories that the search engines have already flagged. They do not pass any value. In fact, submission to a large number of them could be seen as a spam signal, although it is unlikely that any action would be taken solely on this signal alone.

Detecting directories that pass link juice

The process is relatively simple for directories that pass sustainable links, as defined earlier:

- Investigate their editorial policies and see whether they conform to what search engines want.
- Investigate the sites they list. Are they high-quality, valuable resources that do not contain spam or manipulative SEO tactics?
- Investigate their track record. Do they enforce their policy for real? This may be a bit subjective, but if there are lots of junky links in their directory, chances are that the policy is just lip service.
- As another check, search on the directory name and see whether there is any SEO scuttlebutt about the directory.

The process is a bit harder for directories that do not conform to the policies search engines prefer. There are still some things the publisher can do:

- Search on the name of the directory to see whether it shows up in the search engine results. If not, definitely stay away from it.
- Take a unique phrase from the directory's home page and see whether it shows in the search engine results. If not, definitely stay away from it.
- Does the directory have premium sponsorships for higher-level listings? This is a sure signal that indicates to the search engines that the directory's editorial policies may be secondary to its monetization strategy.
- Does the directory promote search engine value instead of traffic? This is another bad signal.
- Evaluate the directory's inbound links. If the directory is obviously engaged in shady link-building tactics, it is a good idea to avoid it.
- Is the directory's target audience webmasters and SEO practitioners? If so, stay away from it.

Incentive-Based Link Requests

Incentive-based link requests use an incentive as part of the process of requesting a link. This can run dangerously close to violating the search engines' Webmaster Guidelines, but there are also ways to do it that remain compliant with those guidelines.

Giveaways

Content syndication can be a powerful tactic, and we will cover this in more detail in "[More Approaches to Content-Based Link Acquisition](#)" on page 314. The concept is to develop content with the intent of publishing it on third-party sites. The site that is publishing the content provides a link back to your site in return for being permitted to publish the content. Ideally, this content should be distinct from the content on your own site to avoid any risk of duplicate content issues.

Widgets are another way of syndicating content to third-party sites. They have the advantage of coming with a way to package the content so that it is not seen as duplicate content. This is because they are usually implemented in JavaScript such that the web page publishing the widget calls back to a remote server to fetch the content. The result is that a search engine crawler does not see the content.

This also results in any links embedded within a widget being invisible to the search engine as well. However, it is possible to implement a widget in such a way that it has an HTML wrapper around it with a simple HTML text link in it, a link that is visible to the crawler.

Popular widgets can get adopted by a large number of websites, and can result in a large number of links as a result. Widget campaigns can also result in links to deep pages on the site.

Dangerous tactics

Of course, the incentives can be over the line too. For example, proactively going out and buying links might be a way to rapidly acquire links, but the search engines are increasingly good at detecting these types of links algorithmically (and discounting or even penalizing them).

Google also provides a method for third parties to report paid links directly to Google on an anonymous basis. There is every incentive for those third parties to do so.

Based on these risks, buying links is a dangerous practice, and it is harder to safely execute than it appears to be on the surface.

Another dangerous tactic is doing a large percentage of your link building through reciprocal links. Once again, this is easy to do in principle. It is not hard to find sites that will accept the "link to me and I will link to you" value proposition.

However, the search engines potentially view this as barter. They are not likely to question a few selected link exchanges with sites closely related to yours. It becomes a problem when the

link swapping becomes a significant portion of your backlink profile. That is a situation that looks manipulative, and they will most likely discount those links.

Direct Link Requests

If the site looking for links is a high-quality site with unique and/or authoritative content, the publisher may simply need to tell other publishers about what it has.

If there is already a relationship between the publisher requesting the link and the publisher being asked to provide the link, this process is pretty easy. The requesting site sends the other party a note with whatever pitch it wants to make. This pitch is easy to personalize, and the nature of what is said is likely to be guided by the existing relationship.

However, if the person requesting the link does not know the person from whom she is requesting the link, it is a very different ballgame. [Table 7-2](#) summarizes how to decide how to contact a site, including how much effort to put in.

TABLE 7-2. Categorizing the value of potential links

Low-value sites	Targets may result from large-scale research. Contact is by email and is personalized but likely in a somewhat automated way. These types of sites are being relied on to obtain a volume of links. No customized content is developed.
Medium-value sites	Targets may result either from large-scale research or from industry knowledge by the publishers of the link destination site. Contact is by email and is personalized, likely done by a human but with only low to moderate levels of effort in customizing a template. No customized content is developed.
High-value sites	Targets are identified by principals of the business or senior marketing people, or through market analysis. Email contact is entirely customized and tailored to the targeted site. Phone calls may also be used in pursuit of these links. Content may be developed just to support a campaign to get links from these types of sites.
Very-high-value sites	Targets are identified by principals of the business or senior marketing people, or through market analysis. Email contact is entirely customized and tailored to the targeted site. Phone calls may be advisable in pursuit of these links. Face-to-face visits may also be involved. Content may be developed just to support a campaign to get links from these types of sites.

Creating a value proposition for direct requests

One of the key points is how the publisher can make its pitch interesting enough for the potential linking site. As noted earlier, this starts with understanding what content or tools the publisher has on its site that would be of interest to the potential linking site.

Sites do not link to other sites for the purpose of helping those sites make money. They link because they perceive that the other sites' users might value the content or tools on their own sites.

This relates to the fundamental structure of the Web, which is designed around the notion of interlinking related documents. Positioning a site or a section of a site as being related to the potential linking site is a requirement of each link-building request.

With high-value sites and very-high-value sites (as defined in [Table 7-2](#)), it may be worth spending a bit of time and energy on defining a compelling value proposition. In many cases, it is even worth doing custom content development to increase the perceived value and relevance of the content to the potential linking site.

In all cases, developing a link request value proposition begins with understanding the nature of the potential linking site's content, and then deciding how to match up the content of the requester's site with it.

Once you have done this you are in a position to contact the site publisher and communicate what you have to offer, and request a link.

Basic email pitch

The most important thing to remember is that the person you are emailing to request a link probably did not wake up this morning wondering what links she was going to add to her site. And certainly, she was not expecting or waiting for your email. Basically, you are interrupting her to ask her to do something for you, and she has no prior reason to trust you. Based on this, there are few simple guidelines you should follow when making a link pitch:

- Keep it simple and short. The person you are contacting is receiving an email that is unsolicited. She is not going to read a two-page email, or even a one-page email.
- Clearly articulate the request. It is an investment to get someone to read an email, and it is critical that the pitch be clear about the desired result.
- Clearly articulate why your site deserves a link. Generally speaking, this involves pointing out the great content or tools on the site, and perhaps citing some major endorsements.
- Follow each and every guideline of the [CAN-SPAM Act](#). Unsolicited emails are not illegal as long as they follow the guidelines of the act. Do not even think about violating them.

Manual Social Media Link Creation

Some social media sites, such as LinkedIn (<http://www.linkedin.com>), allow you to link back to your own sites in your personal profile, and these links are not NoFollowed (i.e., they do pass link juice). Leveraging this can be a great tactic, as it is simple and immediate.

In the case of LinkedIn, the process takes a few steps. There are two major things you need to do:

1. Make sure you list selected web pages in your profile.
2. Go into the Edit Public Profile Settings section and let LinkedIn know you want the web pages you specified to be visible in your public profile.

The preceding process was what was required as of early 2009, but the specifics may evolve over time as LinkedIn releases updates, or LinkedIn may start **NoFollow** those links. Other social media sites that do not **NoFollow** links in their public profiles include Digg, Propeller, Kirtsy, Technorati, MyBlogLog, and Mixx.

Two longer lists of such sites are at <http://www.searchenginepeople.com/blog/22-dofollow-social-media-sites-offering-profile-links.html> and <http://www.seomoz.org/social-directory>. The policies of sites regarding **NoFollow** change on a regular basis, so you should check the current status of the links they provide before investing a lot of time in pursuing this type of strategy.

Another way to create links manually in social media environments is by visiting social media sites, forums, and blogs and leaving behind comments with self-referential links in them. The major steps of this process, using blogs as an example, are as follows:

1. Build a list of blogs that are related to your topic area.
2. Start visiting those blogs and adding comments without linking back to yourself, and develop a relationship with the author(s). The early stages of the relationship begin when the author starts responding to your comments. You can even reach out to the author through one of the major social networks.
3. Once the relationship has been built and seems solid, let the author know about a related value-add resource you have, either through direct contact with her or in a comment. Make sure there is a real connection between your resource and the content from the author (even if you have to create the content on a custom basis, that's OK).

These steps are meant to be conservative to avoid a backlash from the owners and/or the authors of the blog. You can extend this process to forums or social media sites as well.

There are ways to be more aggressive with this. Some publishers do not really care about building relationships first and want to push the process much faster. However, there are two significant issues with this:

- Depending on the level of aggressiveness, it may be a violation of the Webmaster Guidelines and the search engines may choose to take action against the publisher who pursues this.
- There could be a backlash from the community itself. Offending one blogger may not be a huge issue, perhaps, unless she is very influential. Offending hundreds of bloggers would probably be much worse, particularly if you are trying to establish your site as authoritative in a topic area. In forums, blogs, and social media sites, offending people can quickly scale to a problem of large proportions.

Gray Hat/Black Hat

As we previously discussed, some publishers choose to push the limits or ignore the Webmaster Guidelines in their quest for links. On the next few pages we will look at some of the more popular tactics in detail.

Buying links for SEO

One of the more popular techniques is to buy links. This has two significant advantages:

- It is easy. There is no need to sell the quality of the content of your site. The only things that need to happen are determining that the third party is willing to sell a link, and setting a price.
- Since the link is an ad, you can simply specify the anchor text you want. Anchor text is a powerful ranking signal, and this is one of the major reasons people engage in link buying.

Google's policy on paid links. The major downside is that buying links for SEO is against Google's Webmaster Guidelines. Here is a brief summary of Google's policy on paid links:

- Links given in return for compensation should not be obtained for purposes of increasing PageRank.
- The link should be freely given, and the publisher of the potential linking site should be informed of what the publisher is doing. An example of a link where the publisher is not informed is one that is hidden in the `NoScript` tag of a JavaScript-based widget.

Google is not saying that publishers should not be able to buy ads on the Web. Its policy is that links should be purchased only for the traffic and branding value they bring. Google also recommends that publishers selling ads on its site use the `NoFollow` links, which means they will have no SEO value.

On another note, PPC campaigns using AdWords, Yahoo! Search Marketing, and so on are not considered a violation of the policy against paid links. This is because such links are easy for the crawlers to recognize, and due to their paid nature, they do not pass link juice.

Methods for buying links. There are three major methods for buying links. These are:

Direct link advertising purchases

This method involves contacting sites directly and asking them whether they are willing to sell text link ads. Many sites have pages that describe their ad sales policies. However, sites that openly say they sell text links are more likely to get caught in a human review, resulting in the link being disabled from passing PageRank by Google.

Link brokers

As we mentioned earlier, link brokers are companies that specialize in identifying sites selling links and reselling that inventory to publishers looking to buy such links.

The major danger here is that ad brokers may have a template of some sort for their ads, and a spider can recognize a template as being from a particular broker.

Charitable donations

Many sites of prominent institutions request charitable contributions. Some of these provide links to larger donors. Search for pages such as the one at <http://genomics.xprize.org> that links to its supporters. This may be considered a legitimate link if there is an editorial review of which sponsors receive acknowledgment via a link on the institution's site.

Sometimes these types of links do not cost much money. However, Google frowns upon this tactic, so it's best to use it with care. One way a publisher can potentially make the tactic more acceptable is to support causes that are related in a material way to its site. However, it is not clear that Google would find this acceptable either.

Finding these types of sites may seem hard, but the search engines can help with this. One way is to use a series of related searches that will expose sponsor sites. For example, you could go to the search engine of your choice and search on *sponsors*.

This should bring up a number of sites accepting sponsorships. However, you may want to target that search a bit more. If you have a nursing-related site, the next step might be to search on *nursing sponsors*, or even *sponsors inurl:nursing*. You can also try other related words, such as *donors*, *donations*, or *patrons*.

Strategies that are not considered buying links. It is worth noting that in some strategies, money is involved in obtaining a link, yet such links are not considered to have been bought. Here are some examples:

- Using a consultant to help your articles reach the home page of social media sites such as Digg (note, however, that this practice may not meet with the approval of the social media site)
- Paying a PR firm to promote a site
- Paying a link-building firm to ask for (as opposed to buying) links

The key point is that these strategies do not compensate the site itself for the links given, and the links are given freely.

Link farms/link networks

In the early days of search, publishers developed link farms and link networks as tactics for gaining cheap links. A *link farm* is a website or a group of sites whose primary reason for existence is to cross-link between themselves and other websites. Generally speaking, the links are created through aggressive reciprocal linking.

Since these sites are typically very heavily interlinked, they can be pretty easy to detect. Part of the reason is that since they have little redeeming value, they typically do not have

high-value links coming in to them from other sites, and most of the links result from various cross-linking schemes.

Link networks are a similar concept. The network exists for the purposes of creating links between sites, and it can be a bit more sophisticated than a link farm. For example, you could create a club where publishers agree to contribute a link in return for getting a link from somewhere else.

If managed with great care, the clustering of links between sites can be limited, and this can be a bit harder for search engines to detect. However, the tactic remains highly exposed to a disgruntled webmaster simply reporting the scheme to Google.

A related concept is the notion of *three-way link swaps* (a.k.a. triangular link swapping), where Site A links to Site C in return for Site B linking to Site A. In this scenario, Site C may be the site the publisher is trying to promote, and Site B may be a site it uses to provide low-value links to people it trades links with.

This is almost always a scam, because Site B is probably a low-value site with little to recommend it. So, the publisher of Site A is providing a good-quality link in return for a low-quality one.

Automated link dropping

Spam tactics can include the concept of creating a bot that crawls around the Web looking for open forums and blogs and leaving behind automatically generated comments. Clearly this is spam, as no human is involved in the comment process (other than the programmer) and no effort was made to read the blog post or forum where the comment was left.

The great majority of these comments are deleted or `NoFollowed` by the blog or forum software content management system (CMS), but the spammer does not care because she is operating on a large scale. If she leaves behind 1 million comments and 98% of them are filtered by one means or another, she still ends up with 20,000 links.

This is, of course, a very risky tactic. The search engines may be able to detect this behavior algorithmically, or competitors can recognize it and turn you in via a spam report. *We do not recommend this tactic.*

Spammy giveaways

As we previously discussed, you can use widgets as part of a link-building campaign. However, this tactic can be abused as well. One popular way of doing that is to develop a JavaScript-based widget and then embed a link back to the widget publisher's site in the `NoScript` tags. Since this is the code the browser presents to users who do not support JavaScript, and the search engines do not run JavaScript, the crawlers see the link and not the widget.

An example of someone who did this is discussed in the post "Another Paid Links Service Disguised As Hit Counter" (<http://blogoscoped.com/archive/2008-02-06-n19.html>). Four days after

this post went up, the sites referenced lost all of their high rankings in Google, and therefore lost most of their traffic.

However, be aware that making the link visible is not enough to make this practice legitimate in Google's eyes. Google has confirmed that it considers tactics such as the use of unrelated widgets for link building a no-no, even if the link is visible.

The underlying reason for this is that if the link is unrelated to the widget, it is pretty unlikely that the link given actually represents an endorsement of the web page receiving the link. The goal of the person installing the widget is to get the benefit of the widget's contents.

As another example, Google considers the use of so-called sponsored WordPress templates with embedded links, even if they are visible to be spammy as well—unless, of course, the publisher distributing the WordPress template is in the WordPress template business. The key issue that needs to be addressed in these types of link-building campaigns is relevance.

NoFollow uses and scams

One of the most common uses of `NoFollow` is to `NoFollow` all links in comments on a blog or in a forum. This is a common and legitimate use of the tactic, as it is used to prevent spammers from flooding forums and blogs with useless comments that include links to their sites.

However, as you might expect, people did implement some common scams using `NoFollow`. One simple example of this is that someone proposed a link swap and provided the return link but `NoFollowed` it. The goal of the site using the `NoFollow` was simple: to get a link that passes link juice, but not to provide any link juice in return.

However, Google announced that it had changed the way it handles `NoFollow` in June 2009. At this point in time, `NoFollow` still prevents the passing of link juice to the site receiving the link, but it is lost to the linking site as well. As a result, this scam no longer benefits those who attempt to use it.

Choosing the Right Link-Building Strategy

The successful link-building strategy is built on painstaking research and methodical strategizing. You can put together a link-building campaign in many ways, and making the wrong choices can lead to a poor return on your link-building investment. You should also consider which tactic will bring the best long-term value.

Another consideration is the available resources and how easily the link-building process will scale. If you have a site with 10,000 inbound links and your campaign is going to net you 100 new ones, that is not going to move the needle for you (unless all of those links are pointing to an individual page targeting a single keyword search term/phrase, or a smaller number of pages). This is a key point to consider when deciding what type of link-building campaign to pursue.

Outline of a Process

The process for choosing the right link-building strategy is complex because of the number of choices available to you. Nonetheless, a methodical approach can help you determine what the best few choices are for your site. Here is an outline of how to approach it.

Identify types of sites that might link to a site like yours

Some example types of target sites include:

- Noncompeting sites in your market space
- Major media sites
- Bloggers
- Universities and colleges
- Government sites
- Sites that link to your competitors
- Related hobbyist sites

Make sure you take the time to answer the question “why would these sites link to us?” Think broadly here, and don’t limit the possible answers based on your current site. In other words, perhaps you can create some new content that would be compelling to one or more of your target groups.

Find out where your competitors get links

Getting detailed information on who links to your competitors is easy. Simply use tools such as [Yahoo! Site Explorer](#), [Linkscape](#), [Majestic-SEO](#), or [Link Diagnosis](#), each of which will give you a list of the sites that link to your competitors.

Although Yahoo! Site Explorer is a search-engine-provided tool, the data it supplies is much more limited than that of the other tools, so consider using one of the other tools to get a better look at your competitors’ links. The data will be more comprehensive, and it will include highly relevant information on PageRank (and mozRank and mozTrust, in the case of Linkscape).

Once you have that data, look at the most powerful links they have (as measured by PageRank or mozRank/mozTrust) to identify opportunities for your site. Perhaps they have received great links from national media or a set of government sites. By seeing what has worked for them, you can get ideas on what may work for you.

Of course, planning to contact people who link to your competitors is a good idea, but do not limit your link-building strategy to that alone. Contacting websites that link to your competitors may result in your getting links from 10% (in a good campaign) of the people you contact. Chances are that your goal is to beat your competitors, not follow them.

The key focus is to extract data from the competitors' backlinks that helps you decide on your overall link-building strategy. Use this to enhance the list of sites that might possibly link to you.

You can also expand on this concept by looking at "similar pages" to top ranked sites (look for the "Similar" link in the search result for the site) in your keyword markets. Similar pages that keep showing up for different keywords are squarely in the topical link neighborhood. Look at who is linking to them too.

Review your website assets

Now that you have a refined list of targets and a sense of why each group may potentially link to you, review what you have on your site and what you could reasonably add to it. This should include any existing content, new content you could create, tools, or even special promotions (provided that these are truly unique enough and you have enough presence for people to notice and care).

One key aspect of this is that the content needs to be unique and differentiated. Content that can be found on 100 other sites is not going to attract many links. Even if the content is non-duplicate and original, it should have something to offer or say that is differentiated from other content, rather than simply being a rewrite of someone else's article.

The highest-value potential linkers probably know their business and will recognize such simple rewrites, and in any event are going to want to focus their links on unique new content and tools. Content that leverages the publisher's unique expertise and presents a new take on things, or new data, will be far more successful in the link-building process.

Think of your content plan in a business case format. If you were able to create some new block of content at a cost of x dollars, and you think it would provide you with some set of links, how does that compare to another link-building opportunity and the cost of the content (or tools or promotions) required to chase that one?

Ultimately, you will want to build a road map that provides you with a sense of what it would cost to chase a potential group of linkers and the value of each group of linkers. Your chart might look like [Table 7-3](#).

TABLE 7-3. Prioritizing among link-building projects

Cost to pursue	Value
\$15,000	High
\$30,000	Medium
\$10,000	Low
\$30,000	Very high
\$4,000	Low

Once you have this in hand, you can quickly narrow down the list. You'll probably pursue the high-value campaign, and should continue to consider the very-high-value campaign and the low-value campaign that costs only \$4,000 to pursue. The other campaigns just don't seem to have comparable cost versus return.

Identify any strategic limitations

The next step is to outline any limitations you may need to place on the campaigns. For example, if you have a very conservative brand, you may not want to engage in social media campaigns through Digg (which is not a conservative audience).

Identify methods for contacting potential linkers

You must undertake some activities to let potential linkers know about your site. There are two major categories of methods: direct and indirect. Some examples of direct contact methods include:

- Email
- Social media sites, using the messaging features of a social media property to make contact with potential linkers
- Content syndication (contacting people and offering them great article content for placement on their site)
- Blogger networking (building relationships by commenting on others' blogs, writing content intended to be of interest to them, and letting them know about it)

Some examples of indirect methods include:

- Social media campaigns (including Digg, Propeller, Delicious, StumbleUpon, and others)
- PR
- News feeds (through Yahoo! News and Google News)

Link-Building Process Summary

Sorting out where to start can be a difficult process, but it is a very high-return one. Don't just launch into the first campaign that comes to mind, as it can hurt your overall results if you spend six months chasing a link-building plan that does not bring you the right results.

In addition, link building is a form of marketing, and other marketing considerations come into play. Looking at it from the other point of view, cleverly devised link building and marketing campaigns can provide both branding and link-building campaigns. Consider the famous video campaign by [Blendtec](#) of blended iPhones, golf clubs, and so forth, available on YouTube and on its [Will It Blend](#) website.

The Will It Blend site has more than 29,000 links to it from other sites, all entirely natural and earned through editorial recommendations. Not bad!

Putting It All Together

The final step is to consider all these things together to come up with an integrated strategy. You can think of this as having the complete strategic picture in mind as you approach link building. At this point, final decisions are made about the content strategy and which link-building targets to focus on.

Execute aggressively

A world-class link-building campaign is always a large effort. The resources need to be lined up and focused on the goal. Publishers nearly always experience bumps in the road and difficulties. But it is critical to line up the resources and go for it. It is also critical to be persistent.

It is amazing how poorly many companies execute their strategies. Publishers that execute aggressively inevitably gain an edge over many of their competitors. Of course, if other competitors also focus heavily on link building, it becomes even more important to push hard.

Conduct regular strategic reviews

Link-building strategies should evolve in the normal course of business. As the implementation moves forward, lessons are learned, and this information can be fed back into the process.

New strategies are also conceived over time, and some of these are great ones to pursue. Also, sometimes the initial strategy goes great for a while, but it begins to run out of steam. Publishers should have a constant stream of ideas that they are feeding into their link-building plans.

Create a link-building culture

Publishers should also train many people within the organization about their link-building plan, its goals, and how it will help the business. The purpose of this is to engage the creativity of multiple team members in feeding the stream of link-building ideas.

The more ideas you have, the better off you'll be. The quality of a link-building campaign is directly proportional to the quality of the ideas that are driving it.

Never stop

Link building is not something you do once, or once in a while. In today's culture, the search engine plays an increasingly large role in the well-being of a business, and at least for the moment, linking relationships greatly determine the fate of sites on the Internet. Consider the business that implements a great link-building campaign, gets to where it wants to be, and then stops.

What happens to the business when its competitors just keep on going? It gets passed and left behind. Publishers who are lucky enough to get in front need to be prepared to fight to stay there. Publishers who are not out in front should fight to get there.

More Approaches to Content-Based Link Acquisition

Content is your most important asset in link building. Leveraging your content—together with your users and online networks—can lead to scalable link acquisition and some exciting results.

A Closer Look at Content Syndication

The concept in content syndication is to develop content with the intent of publishing it on someone else's site. In return for providing the content, the author gets a link back to her site. This is a legitimate strategy in the eyes of the search engines because the site is endorsing the content by accepting it, and the return links are an acknowledgment of that endorsement.

It is also often possible to get targeted anchor text in this scenario. Be careful, however, because if the anchor text is unrelated to the article itself, it will not comply with what the search engines want publishers to do for link-building techniques.

There are a couple of important points to watch for when syndicating content:

- The publisher should strive not to distribute articles that are published in the same form on the publisher's own site. Search engines will see this as duplicate content. Although the search engine's goal is to recognize the original author of a piece of content, it is a difficult job to do perfectly, and it does happen that the search engines make mistakes.

When looking to distribute content published on a site, the best practice is to rewrite the article, add some things to make it a bit different in structure and in its material points, and syndicate that version of the article. One way to do this is to take a different angle on the discussion within the content. Therefore, if the site publishing the syndicated article ranks high for key search terms, it is not a problem for the author's site.

If you must syndicate the exact article you have on your site, have the article link back to the original source article, not just your site's home page. That will serve as a signal to help the search engines identify which version is the original.

- When considering the syndication of content, it makes sense to study the site's content needs and then custom-tailor the content to those needs. This practice maximizes the chances of the target site accepting the article.

One variant of content syndication is to generate articles and then submit them to *article directories*. Many of these types of sites exist, and many of them are, frankly, pretty trashy. But there are good ones as well. The distinction between the trashy and good ones is relatively easy to recognize, based on the quality of the articles they have published. Many of these article directories allow publishers to include links, and these links are followed. [Table 7-4](#) lists some that are worth considering

TABLE 7-4. List of quality article directories

Directory site	Category
http://www.buzzle.com/	General
http://www.ezinearticles.com	General
http://www.work.com	Business
http://www.articleblast.com/	General
http://www.articlecity.com/	General
http://www.articlesbase.com	General
http://www.mastersmba.com	General
http://www.trans4mind.com/	Self help
http://www.pandecta.com	e-business
http://www.promotenews.com/	Web marketing
http://www.a1articles.com/	General
http://www.articleclick.com	General
http://www.articlesphere.com	General
http://www.bestmanagementarticles.com	Business
http://www.buildyourownbusiness.biz	Business
http://www.ibizresources.com/	Business
http://www.isnare.com/	General
http://www.redsofts.com/articles/	General
http://www.sbinformer.com/	Small business
http://www.valuablecontent.com	General

Note that even if these do appear to pass link juice, some of them may be disabled on the back end by Google, but there is no way to tell which ones. For that reason, it is best to focus on whether the article directories themselves appear to be of value. This maximizes the chances that the search engines will value them as well.

In general, though, search engines are not going to punish a publisher for putting an article in these directories, and many of them may provide somewhat helpful links. However, this is not a high-value activity, so although you can pursue it, you should probably not make it the focus of your link-building efforts.

Leveraging User-Generated Content

Providing users with ways to contribute content directly to your site can be an effective tactic. There are many ways to do this:

Open up a forum on your site

One of the biggest challenges with this strategy is achieving critical mass so that active discussions are taking place on the site. This typically requires a substantial amount of traffic to accomplish, but in the right situations it can become an activity for developing interesting content with little effort.

Launching a blog and inviting third-party contributors

One of the best ways to do this is to contact respected members of your market space and see whether they would be willing to make written contributions to your blog. They may do this simply because they would like the exposure, or you can pay them for doing it.

More selectively inviting third-party contributions

A blog platform may be more than you want to do, but you can still ask respected members of your community to consider contributing articles to your site.

Of course, the contributed content does not need to be an article or a post. You can seek out photos, videos, cool new tools—anything that may be of interest to users.

With each of these strategies, one of the big questions is whether the method for contributing content is open, strictly controlled, or somewhere in between. In other words, can any user come along and post a comment in your forum? Or do all users have to have an editorial review first? Editorial reviews may significantly reduce spam attacks, but they are a barrier to the development of active discussions.

In the case of forums, engaging discussions can attract links. For an example from the world of SEO, [Search Engine Roundtable](#) is a frequent linker to discussions in the [WebmasterWorld Forums](#). The key to this is the critical mass of discussions that take place on these forums.

The reason the tactics involving third-party authorship can result in links is that most people have pride in what they have created and want to show it off. As a result, they will have a tendency to link to their content from other sites where they contribute, or their own website.

It is a good idea to make this easy for them. Provide sample HTML that they can use, or badges that confer the value of recognition—for example, a badge that says something like “Valued Yourdomain.com Contributor,” or “Contributing Offer.” Only people who are authorized contributors are allowed to post such a badge, so it becomes an honor to be able to do so.

Creating Link Bait/Viral Content

Link bait is the term that some use to refer to the notion of creating content for the specific purpose of acquiring links. The content is published on your own site or perhaps on another website, and it is compelling enough that lots of people link to it. Such content can take a

couple of forms. For example, it can be content that is designed to provide enough additional value that people will want to reference it.

Other popular methods are doing something controversial, something funny, or something that simply draws a strong emotional reaction. Each market space has some hot buttons, and these buttons can be pushed with an opinionated article.

Coming up with link bait ideas

It is easy to come up with content that people will want to link to, but it does take some effort to come up with the right kinds of ideas. Here is a four-step process for coming up with and picking ideas:

1. Write down everything.

Just collect any and all ideas that come to mind. Do not censor or edit yourself during this phase; take down every idea no matter how bizarre, idiotic, or farfetched it may sound.

2. Break down your ideas.

Once you've squeezed every last drop of creative juice from your head, it is time to filter everything out. It is a good idea to break down each idea (no censoring or editing yet) into its Concept and Content components. That is, what is the format (Concept) for the suggested link bait (tool, widget, top 10 list, how-to guide, blog post, etc.) versus what is the subject (Content) of the suggested link bait (Wii, iPod, PPC ads, pigs, celebrity weddings, etc.)? Separate these into two lists.

3. Evaluate the content.

Ignoring your Concept list, critically evaluate your Content list. Are some ideas time-sensitive? Can some wait for a relevant piece of news to complement them? Are there ideas you'd really like to write about? Are there ideas that can go into storage for a dry spell?

4. Mix and match.

Once you've prioritized the content, you can mix and match it with your Concept list. No story/content is beholden to the original format in which you brainstormed it. Is your story something that might make Digg? Then consider the concepts that do well there: top 10, how-to, and so on. Can it be interactive? Perhaps a tool or poll concept would be effective. By marrying your priority content to the most appropriate concept, you can optimize the effectiveness, reach, and novelty to your intended audience.

If you use this process or something similar, you'll probably notice that you quickly generate a handy repertoire of concepts. Once you have these down, you can turn virtually any random idea that pops into your head into link bait.

How far should you go with link bait?

Most folks in the corporate communications, PR, and legal departments shy away from anything potentially controversial, and for good reason, right? But then, why would a company selling life insurance online dare to venture into the taboo topic of weird and disturbing death trivia? Sounds crazy, doesn't it?

But that's exactly what [Lifeinsure.com](http://www.lifeinsure.com/information/19-things-about-death.asp) did with its link bait article, "The 19 Things You Probably Didn't Know About Death" (<http://www.lifeinsure.com/information/19-things-about-death.asp>). With such goodies as "After being decapitated, the average person remains conscious for an additional 15–20 seconds," you can imagine how much of a hit the article was with the irreverent alpha geeks that make up the Digg community. The article made it to the Digg front page, which in turn got it in front of countless bloggers and social bookmarkers.

Surely the success of this article in attracting links has contributed to Lifeinsure.com's impressive Page 1 ranking for *life insurance* that it maintained for many months. Not surprisingly, though, this contentious article is nowhere to be found in Lifeinsure.com's navigation hierarchy, so customers and prospects are unlikely to ever stumble across it.

Potential linkers also love a good corporate citizen, so be one. Consider such activities not as an expense, but as an investment that will generate a return in the form of links. With [Second Chance Trees](#), social media marketing agency Converseon created a charitable initiative using internal resources and expertise that could have otherwise been utilized for billable work.

The idea was to create an island in Second Life where players could purchase a virtual tree with Linden dollars and plant it. This would then trigger the planting of a real tree of the same species in an ecologically sensitive region, such as a Central or South American rain forest. For a charitable endeavor, the payoff was huge. High-value links came from news outlets, the blogosphere, organizations, and elsewhere.

Do not be afraid to be bold or off the wall. You do not always have to toe the corporate line. If you're thinking that this will garner links that aren't very relevant to your business and industry, you're probably right. But remember that one component of PageRank is topic independence. Tests we performed show that high-PageRank-endowed yet topically irrelevant links still help, and they can help a lot. Definitely still work to acquire topically relevant links as well, but do not neglect the off-topic ones too.

Encourage link bait to spread virally

You can extend the notion of link bait distribution by creating something you can pass around. For example, a hilarious video clip might be passed around via email. Provided you make it easy for people to determine the video creator (presumably your company) and to visit your site, this type of campaign can garner a lot of links. Be aware, though, that if you host the video on a video-sharing site such as YouTube, most people will link to YouTube and not to your site.

Incentive-Based Link Marketing

Providing an incentive for people to provide a link back to you can work well. Of course, you need to do this with care, because there are incentives that would be indistinguishable from the outright purchasing of links and therefore risky. However, you can use certain tactics in which the resulting links will likely still have editorial value. This includes creating badges or levels of recognition that users will want to show off, as we discussed in “[Leveraging User-Generated Content](#)” on page 316. Here are a couple of other ideas.

Helping Other Sites Boost Their Value

You can get people to link to you because linking to you can enhance their site’s value. One example of this is a program that gives sites awards for excellence of some kind. This is particularly effective if your site is a respected authority in your space.

If that is the case, you can consider picking out the top 10 sites related to your space. Then you can send these sites an award badge with a link to a review that lives on your site.

Another example is various forms of certification. For example, Google has badges for its AdWords Qualified Professionals and AdWords Qualified Companies, with links back to Google.

Customer Discounts/Incentives

Publishers can offer visitors from certain websites a discount on the publisher’s product. Sites that want to offer the discount to their users simply need to link back to the publisher’s site.

However, Google also sees this as a questionable practice. Although it would seem that the discount offer would be interesting to the third-party site only if it valued the product (or thought its visitors would), Google’s Matt Cutts clearly indicated that Google did not want to value such links in an interview at <http://www.stonetemple.com/articles/interview-matt-cutts-061608.shtml>.

How Search Engines Fight Link Spam

Throughout this chapter, we have provided many examples of how spammers try to circumvent search engine guidelines to obtain high rankings for sites that may not deserve those rankings. Of course, the search engines do many things to fight link spam.

Algorithmic Approaches to Fighting Link Spam

The major approach the search engines use is to design algorithms that can detect and act on link spam. There are a number of things they can look at algorithmically. Here are a representative few:

Links labeled as advertisements

The search engines can scan for nearby text, such as “Advertisement,” “Sponsors,” “Our Partners,” and so on.

Sitewide links

Sitewide linking is unnatural and should be a rare part of your link mix (purchased or not). The only exception to this is the interlinking of all the sites owned by your company, but this presumes that the search engine will understand that all of your sites are from the same company. In general, sitewide links are a serious flag, especially if you have a lot of different sites that do this for you, or if a large percentage of your links are sitewide.

Links sold by a link broker

Of course, link brokers are knowledgeable about the link detection methods listed here, and they do their best to avoid detection with the links they sell. But they can still run into problems. For example, Google took action against a long-time proponent of paid links, [We Build Pages](http://www.webuildpages.com/blog/link-techniques/paid-links-arent-worth-it/), resulting in We Build Pages changing its stance on the subject (<http://www.webuildpages.com/blog/link-techniques/paid-links-arent-worth-it/>).

Selling site has information on how to buy a text link ad

Search engines can detect sites that provide information on how to advertise with them. Combined with other clues about links being sold on the site, this could lead to a review of the site selling the ads and a discounting of the links.

Relevance of your link

It is a powerful clue if your link is not really that relevant to the page or site it is on.

Quality of neighboring links

Another clue would be the presence of your link among a group of links that are not tightly themed.

Location outside main content

The search engine can detect when your link is not part of the main content of the page. For example, it appears in the left or right column of a three-column site, and the main content is in the middle.

Perhaps you can avoid all of these pitfalls, but one more problem remains: people can see that you are buying links and choose to report your site, using Google’s authenticated paid links reporting form (<https://www.google.com/webmasters/tools/spamreport?hl=en&pli=1>). Note that you need to be logged in to a Google account to see the form. Here are some examples of people who might take this action:

Someone reports your site for buying links or for some other reason

Who would do this? Your competitor! If your competitor submits an authenticated spam report to Google, Google will look at it and act on it.

Someone reports the site you bought links from for selling links or for some other reason

A competitor of yours can do this, or a competitor of the site selling links can do this. Once a search engine figures out that a site is selling links, it is possible that this could trigger a deeper review of the sites that were buying those links.

A disgruntled employee leaves your company, the broker, or the site you bought links from and reports your site

For decades, many companies have had a practice of escorting fired (or laid off) employees out of the building. The reason for this approach is that people get upset when they lose their job. However, this practice would not prevent such a person from reporting your site in a spam report to a search engine. Even though that may be a violation of the confidentiality agreement you probably have with your employees, you would never know, because there is no transparency in spam reporting.

A search engine employee does a manual human review of your site

The search engines do maintain staffs of people who conduct human reviews of sites, which they use to proactively find and report spam.

Certainly your competitor reporting your site would be the most likely scenario, but you should not entirely discount the other scenarios.

Other Search Engine Courses of Action

In the case of Google, it is known that one of the basic policies is to punish a site that sells text links by eliminating that site's ability to pass link juice. This is essentially a first course of action. Once this is done, Google could look more closely at the selling site and the purchasing sites for other signs of spammy behavior.

Google has also at times taken the step of altering a site's visible PageRank (the PageRank shown on the Google Toolbar) if it believes a site is selling links. This has been applied to some very significant sites, such as [NewsDay.com](#).

Google can also choose to take action against the site purchasing the links or participating in manipulative link practices. If a large percentage of the site's links are suddenly disabled, this could have a very significant impact on rankings.

Social Networking for Links

When you add social media to the equation, the network effect can multiply the yield from your link-building efforts. You can use this effect to help your content spread virally, or to develop relationships with critical influencers.

Blogging for Links

Blogging can also be effective in link development. How effective a blog will be depends highly on the content on it, the market space, and how the publisher promotes it. The first thing to realize when starting a blog is that it is a serious commitment. No blog will succeed if it does not publish content on a regular basis.

How frequently a blog needs to publish depends on its topic matter. For some blogs one post per week is enough. For others, it really needs to be two to three times per week, or even more often.

Blogging is very much about reputation building as well. Quality content and/or very novel content is a key to success. However, when that first blog post goes up, the blog will not yet be well known, will not likely have many readers, and those that do come by will be less likely to link to a little-known blog.

In short, starting a new blog for the purpose of obtaining links is a process that can take a long time. But it can be a very effective tool for link building. Just remember that patience and persistence are required.

One of the best places to get links to a blog is from other blogs. This is best done by targeting relationships with major bloggers and earning their trust and respect. Here are a few key things to think about when building these relationships:

- Be patient when developing relationships with other bloggers. Trust and respect do not come overnight. And they certainly do not result from starting the relationship with a request for a link.
- The publisher should target a portion of its content at the interests of the major bloggers.
- Over time, this process should turn into links to the publisher's blog from the major bloggers.
- Other lesser-known bloggers will begin to see links on the major blogs and will begin to follow suit.

This can pay big benefits. Vizu, a market research company, published a study (<http://answers.vizu.com/solutions/pr/press-release/20070305.htm>) that showed that 67.3% of people found what blogs to read by following links from other blogs.

It is also important to leverage the social nature of the blogosphere. Publishers just launching a blog should try to provide a personalized response to every person who comments on their blog. One effective way to do this is to send each and every one of them a personalized response by email that shows that the comment was read.

This helps to deepen the interest of the commenter, creates a feeling of personal connection, and increases the chance that the commenter will return, and possibly add more comments. Nurturing the dialog on a blog in this fashion helps that dialog grow faster.

Leveraging Social News and Tagging Sites

Social media sites such as [Digg](#), [Reddit](#), [StumbleUpon](#), [Delicious](#), and others can play a big role in a link-building campaign.

Becoming “popular” on these sites can bring in a tremendous amount of traffic and links.

Although social news sites such as Digg bring lots of traffic, this traffic is usually of low quality, and will have a very low revenue impact on the site receiving it. The real ballgame is to get the links that result.

For example, stories that make it to the Digg home page can receive tens of thousands of visitors and hundreds of links. Although many of these links are transient in nature, a significant number of high-quality links result as well.

When pursuing social media strategies, there are a variety of best practices to keep in mind:

- The number one guideline is to remember that the focus of this discussion is on “social” media sites. Your success on these sites is wholly dependent on the response of the group of people as a whole to your strategy. If you do something that irritates the community, you can quickly find yourself called out and exposed in that community.

Great care is recommended when working with these communities. Make sure you are a positive contributor. Take your time and build a reputation with the community. “Give more than you get” is a great principle to keep in mind.

- If you create articles for social news sites such as Digg, Reddit, and Propeller, focus on topics that relate to keywords relevant to your business. As a result of doing this, your articles can quickly end up ranking very well for those keywords, and the relevance of the inbound links will be high.

The key insight into how to make that happen is to use the competitive keyword in the title of the article itself *and* in the title of the submission of the article to the social news site. These are the two most common elements that people linking to such articles grab when selecting the anchor text they use.

- Sites such as Delicious and StumbleUpon are different in structure. Delicious is a tagging (or bookmarking) site used by people to mark pages on the Web that they want to be able to find easily later. StumbleUpon shares some similarities with Delicious but also offers a content discovery aspect to its service.

Both of these sites have “popular” pages for content that is currently hot on their sites. Getting on those pages can bring lots of traffic to a site. This traffic is of higher quality than what you get from social news sites.

Users coming to a site via tagging sites are genuinely interested in the topic in a deeper way than users who spotted a snappy article title on a social news site. So, although the traffic may be quite a bit lower than on the social news sites, publishers can also earn some quality links in the process.

Tagging sites are best used in attempting to reach and develop credit with major influencers in a market space. Some of these may link to the publisher, and these are potentially significant links.

Forum and Social Network Participation

Building relationships with other bloggers as outlined so far is good, but there are additional ways to interact with others and in the process to let them know you have content they might be interested in. Any environment in which social interactions occur is another good place to invest time and effort.

Sites such as those of the current market leaders—[LinkedIn](#), [MySpace](#), [Facebook](#), and [Twitter](#)—can be used for link-building purposes without actually getting the link juice directly from one of these sites. Major forums that relate to your area of interest also represent great targets.

The guidelines are basically the same: become an active part of the community. The goal is to become recognized as an active and trusted contributor to the community so that based on your contributions there, people will begin to develop an interest in the other things you may have to say.

For example, people may begin asking for your opinions on other related topics. This creates opportunities to take the dialog further. Relationships and networking are essential to increasing readership.

Of course, in the process of building relationships in a social media environment, you also have the potential for people to submit your content into those social media sites, which can lead to links. You can also take the next step and reach out through the social networks or forums to make direct contact with people to let them know about your content. This is similar to emailing people, but with a few important distinctions:

- Publishers can send out communications to their friends on those networks. Assuming that they have treated this designation (that of friend) with any level of seriousness, instead of “friending” everybody in sight, the communication can be a bit more informal than an unsolicited email would be.
- Publishers can also join groups on these networks related to their market space, and then send messages out to those groups. These groups will provide them with the ability to reach new people with related interests.
- Messages broadcast through these networks cannot be personalized, so a more general message needs to be tailored for these types of broadcasts.
- These are social networks. Beware of broadcasting too many messages or poorly targeted messages. Many publishers made this mistake, became pariahs in their communities, and lost the leverage these communities bring to the process.
- Personalized messages can be sent on a 1:1 basis as well.

One strategy for approaching an authority site is to make initial contact by friending someone senior who works at the company that publishes the site. Then the publisher can develop the relationship with that senior person without asking anything of her. Once the relationship is established, a more informal approach can be used to introduce the great content on the publisher's site.

Offline Relationship Building

Leveraging online platforms for the purpose of building relationships makes great sense, but there is no reason to stop there. Do the major influencers in your space speak at conferences? If so, go to one of these conferences and introduce yourself. This can include not only bloggers, but also other people influential in your space.

Another related tactic to consider is contacting the publisher of an authoritative site and offering the publisher a free seminar/webinar, with you as the speaker. Or you could propose a joint marketing campaign with the publisher. Either way, call it part of your company's outreach campaign to build relationships with leaders in the space.

If you do this, make sure you articulate well the unique nature of what you will present. You have to attract their interest with the pitch before you can take the next step. Make sure you bring a lot of value in the actual presentation.

Then collect business cards, answer questions, and make yourself available to answer follow-up questions by phone or email. Once you have done that, you will have a ton of relationships with people involved in the authoritative site.

There are other ways to extend this too. For example, you can sponsor the organization in some fashion. There are, of course, sites that link to their sponsors, and this may be a win, but it is a link that Google will want to discount (because it is "paid"), so you should not count on that aspect of it. The other win is that sponsors often are able to establish a deeper relationship with the organizations they sponsor.

Last but not least, most likely you have other businesses/organizations that you interact with in the normal course of your business. Once again, you can ask them directly for links, but you can also ask them for introductions to other people in your space. This type of networking can be an effective way to build relationships that eventually lead to high-value links.

Some Success Stories Using YouTube

It may become more important for your brand or company to be on YouTube than to be advertised on TV. For some that day has already arrived. As of January 2009, Hitwise and comScore both rated YouTube as the #2 search engine on the Web.

YouTube has launched careers, such as that of YouTuber "Brookers" (<http://www.youtube.com/profile?user=Brookers>), who was hired by Hollywood celebrity Carson Daly because of her zany

videos. YouTube has also brought international fame to previously unknown bands, such as Sick Puppies, a band popularized by the hugely well-liked and inspiring Free Hugs video set to the Sick Puppies song “All the Same” (http://www.youtube.com/watch?v=vr3x_RRJdd4).

Then there are the hugely successful viral campaigns by commercial organizations, such as Blendtec’s [Will It Blend](#) that we referenced previously in this chapter. This is a brilliant video series on various household objects that are run through a Blendtec blender—including marbles, rake handles, and even iPods.

Blendtec isn’t the only company that has had success with YouTube. Intuit, maker of the Quicken, QuickBooks, and TurboTax software, ran a YouTube campaign known as the Tax Rap. It was a pretty off-the-wall idea suggested in a brainstorming session; as luck would have it, Intuit was able to secure rapper Vanilla Ice as its front man. After that, it decided to just pull out all the stops—avoiding any corporate marketing feel to the campaign.

Seth Greenberg, group manager of Online Advertising and Internet Media at Intuit, says, “Rather than people making fun of our campaign we wanted to poke fun at ourselves.” They went to Vanilla Ice’s house in Palm Beach, Florida, and spent several hours there shooting. Vanilla Ice has been a big supporter of the campaign, according to Seth. What is interesting is that Vanilla Ice is a polarizing figure. But that is what is making it a phenomenon on YouTube. ([AdCritic.com](#) said, “I don’t know how to read this campaign....”)

The campaign received more buzz offline than online. It was covered by news outlets such as CNN, as well as local stations. *Entertainment Weekly* listed Vanilla Ice’s Tax Rap as #10 on its Hit List. And it made it onto Page Six of the *New York Post*.

The key to the Tax Rap video campaign (<http://turbotax.intuit.com/taxrap/>) was not just that Vanilla Ice was the front man, but that it also encouraged participation and viewer support. There was a contest with prize money of \$50,000, with users encouraged to create their own rap about taxes to compete for the prize.

Intuit made a sizeable investment, including buying a Contest channel and a Branded channel on YouTube, as well as paying for visibility on the [YouTube.com](#) home page. Those were crucial factors for Intuit’s getting more than 1 million views of its video.

Online jewelry retailer [Ice.com](#) made its first foray into YouTube marketing with its Mr. Cupid interviews of passersby (<http://www.youtube.com/watch?v=yPkW1gnhoQg>). Executive VP of Marketing and Founder Pinny Gniwisch put some videos up of himself conducting impromptu interviews on the streets of New York City, in Times Square, on the ski slopes of Utah, and elsewhere, prior to Valentine’s Day. Pinny said the videos did very well for the company.

YouTube has been used effectively for brand damage control as well. For example, the CEO and founder of JetBlue Airlines put up an apology video (http://www.youtube.com/watch?v=-r_Plq7EAUw) on YouTube because of a Valentine’s Day winter storm incident—a campaign that was well received.

One product that got some excellent brand recognition and building from being on YouTube was Smirnoff's Raw Tea. Smirnoff produced an uproarious music video called "Tea Partay" (<http://www.youtube.com/watch?v=PTU2He2Blc0>), with preppies rapping.

Social media expert Neil Patel (<http://www.pronetadvertising.com/about/>) notes that the problem with most popular YouTube promotions is that YouTube gets the links and the original site usually does not. That means the search engine visibility benefits do not usually transfer to the company's website.

Even without creating your own site or hosting your videos on your main site to draw links, YouTube offers much in the way of brand visibility when the campaign is well executed. That doesn't just mean posting a great video; marketers must also know how to take advantage of the social nature of the site, to build up friends and to get on user subscription lists.

Jonathan Mendez of [Optimize and Prophesize](#) is an evangelist for the power of tags for marketing on YouTube. His advice is to make copious use of tags on your videos (ensuring, of course, that the tags are relevant to the content), to spread your tags out among your clips, to use adjectives to make your videos more visible to folks who are searching based on their mood, to have some category descriptor tags (bearing in mind that YouTube's default search settings are Videos, Relevance, and All Categories), to match your title and description with your most important tags, and not to use natural language phrases or waste tag space on words such as *and* and *to*.

Do not be afraid to make a start, even if it is modest and has little budget behind it. You won't get anywhere without experimenting with the medium. Ongoing experimentation provides you with the best chance to find a formula that works for you.

Social Media Tips for More Sites

Although we have outlined some of the basics of how to work with social media sites, each site has its own quirks and opportunities. What follows is an outline of tips to use on many of these sites.

Wikipedia

You may not think of Wikipedia as being useful in link building since it NoFollows all its external links. It is not a wise idea to simply go onto the site and create a page for your company. Unless your company is well known, it is likely to be promptly removed. However, links from Wikipedia can be valuable because many people treat it like an authoritative site, and if they see your link on Wikipedia they may choose to link to you. This can include some pretty influential people.

Therefore, you may want to build trusted relationships within the Wikipedia community. This will require an investment in time, but it can bring some nice rewards. Here are some specific tips on how to proceed:

- Build up your credibility (long and virtuous contribution history, user profile page with Barnstar awards) before doing anything that could be construed as self-serving. It is not good enough to be altruistic on Wikipedia unless you *demonstrate* it. It has to be visible as a track record (e.g., do squash spam and fix typos and add valuable content, but do not do it anonymously).
- Negotiate with an article’s “owner” (the main editor who polices the article) before making an edit to an article to get her blessing first.
- Use Wikipedia’s “Watch” function to monitor your articles of interest. Better yet, use a tool that emails you (e.g., TrackEngine, URLyWarning, or ChangeDetect).
- The flow of PageRank can be directed internally within Wikipedia with Disambiguation Pages, Redirects, and Categories.
- Make friends. They will be invaluable in times of trouble, such as if an article you care about gets an “Article for Deletion” nomination.
- Do not edit anonymously from the office. This could come back to haunt you. Tools exist that could embarrass you if you do so. One public domain tool, WikiScanner, is able to programmatically take anonymous Wikipedia posts and identify the organization that created them. The tool cross-references the IP address to make the edits with the blocks of IP addresses of more than 180,000 organizations. Do not take the risk.

Wikis

Plenty of other wikis are a lot more edit-friendly than Wikipedia and let you contribute valuable content, get links, and build relationships. Examples include [ShopWiki](#), [The NewPR Wiki](#), [WordPress Codex](#), and conference wikis such as Web20Expo. Some even pass link juice, which is a nice bonus.

Flickr

As of early 2009, Flickr is the most popular photo-sharing site on the Web. It can be an effective tool in gathering traffic and exposure for your website. Lots of people search on Flickr, and images in Flickr can rank prominently in web search results. If people find your images on Flickr, it can cause them to click through to visit your site. As with other social media properties, this traffic can lead to links to your main site, but it is an indirect approach. Here are some key tips:

- Always use tags—as many as possible while still being accurate. Surround multiple-word tags with quotation marks.
- Make descriptive titles for your photos.
- Create thematic sets for your photos.

- If the photo is location-specific, go into Flickr's tools and geotag the picture by finding its location on Yahoo! Maps and then dragging the picture onto the map to pinpoint its location.
- License your photo with Creative Commons and state how you want the user to credit you in your photo's description.

Meetup.com

[Meetup.com](#) is a site designed to help groups of people with interests in a common topic “meet up.” Completely aside from any SEO concerns, these meet-ups can be excellent networking events.

Get involved with relevant local meet-ups and get your [Meetup.com](#) member profile page linked from the Meetups page, which will pass link juice to your profile and then on to your site.

Twitter

Twitter has established itself as the leading microblogging site. It allows its members to contribute microblog posts (a.k.a. “tweets”) that are limited to 140 characters. It has become an environment for real-time communication with a broader network of people. You can use Twitter as an effective platform for promoting your business. It is just another channel for communicating with your customers and market.

The basic concept is to become an active member of the community and build a large network of followers. As with other networking sites, many important influencers spend time on Twitter. If you can use Twitter to develop relationships with these people, that can provide some very high-quality links.

In addition, if you create some high-quality content, you can gain a lot of exposure as a result of getting a lot of “re-tweets” (whereby people forward your message to all of their followers). Particularly interesting tweets can get significant visibility for you and your company. Here are some more specific tips for leveraging Twitter for relationships and links:

- Create a subpage or microsite dedicated to Twitter. For example, Zappos created <http://twitter.zappos.com>. Our recommendation leans toward creating a subpage (e.g., <http://www.yourdomain.com/twitter>) as this is likely to do a better job of passing link juice back into the rest of your site.
- Avoid getting your message blocked by a recipient's email spam filter or adding to an already overflowing inbox by using Twitter's direct messages.
- Contact influencers in your Twitter network by working through those in common with you. Identify the common “friends” with [Tweetwheel.com](#). You can send your request for them (e.g., to check out your latest post) as a direct message.

Social Media Summary

To sum up the social media strategy, it is useful to think about it as an old-fashioned PR/marketing strategy. Marketers used to think about the number of “impressions” they were able to create in people’s minds. The more impressions you can create the better. When TV, print, and radio were the only media to worry about, this was relatively simple to execute.

The current environment is much more complex. The people you are trying to reach use many venues, and all of the various social media properties are part of that picture. Whenever you have a social media property that a large number of people use, there is an opportunity to reach them. Success in reaching them depends on becoming a trusted member of that community.

Since there are many large communities, it can be complex and expensive to establish a presence in all of them. But presence in each one does add to the number of opportunities you have for creating impressions in your target audience. Since major influencers in your market may be using these communities, you have an opportunity to reach them as well.

Pursuing social media sites can be a very effective strategy. It does require a real investment to do well, but when it’s done well the strategy can provide some nice benefits.

Conclusion

In this chapter’s introduction, we noted that “links are the main determinants of ranking behavior.” If you do an okay job on link building and your competitor does a great job, your competitor’s business will grow at your expense. As a result, link acquisition strategies are an essential part of an effective SEO effort.

You should also see link building as an ongoing activity. Each of us has seen cases where a brief focus on link accumulation brought returns that were squandered by abandoning the strategy. Unfortunately for these sites, they lost momentum and rankings to their competitors (the same ones they passed when they were actively building links), and it proved very difficult to catch up to them again.

Link building is not fundamentally different from PR work. Your goal is to acquire positive citations across the Web. The big difference is the technical aspects—focusing on the quality of the referring source, the keywords in the link, and the page(s) to which they point.

People will not link to low-quality content, or sites that offer a poor user experience (unless they are compensated for the link). And unless you are fortunate enough to possess a major brand, people won't link to purely commercial sites either. You have to offer something of value to users, but you also need to offer something unique. Certain content naturally attracts links because it triggers psychological and emotional responses—pride, sharing, newsworthiness, and so on. Leverage these triggers and create a compelling reason for visitors who can influence web content (writers, publishers, bloggers, etc.) to reference your work, and your link growth efforts will be a success.

Great link building comes from a simple idea: "Build great stuff, tell everyone about it, and motivate them to share."

CHAPTER EIGHT

Optimizing for Vertical Search

VERTICAL SEARCH ENGINES FOCUS ON SPECIFIC NICHES OF WEB CONTENT, including images, videos, news, travel, and people. Such engines exist to provide value to their user base in ways that go beyond what traditional web search engines provide.

One area where vertical search engines can excel in comparison to their more general web search counterparts is in providing more relevant results in their specific category. They may accomplish this by any number of means, including making assumptions about user intent based on their vertical nature (an option that full web search engines do not normally have), specialized crawls, more human review, and the ability to leverage specialized databases of information (potentially including databases not available online).

There is a lot of opportunity in vertical search. SEO professionals need to seriously consider what potential benefits vertical search areas can provide to their websites. Of course, there are significant differences in how you optimize for vertical search engines, and that is part of what we will explore in this chapter.

The Opportunities in Vertical Search

Vertical search has been around for almost as long as the major search engines have been in existence. Some of the first vertical search engines were for image search, news group search, and news search, but many other vertical search properties have emerged since then, both from the major search engines and from third parties.

This chapter will focus on strategies for optimizing your website for the vertical search offerings from Google, Yahoo!, and Bing. We will also spend some time on YouTube, which in January 2009 became the second largest search engine on the Web. First we will look at the data for how vertical search volume compares to regular web search. The data in [Table 8-1](#) comes from Hitwise, and shows the top 20 Google domains as of May 2006, which is one year before the advent of Universal Search.

TABLE 8-1. Most popular Google properties, May 2006

Rank	Name	Share
1	Google	79.98%
2	Google Image Search	9.54%
3	Google Mail	5.51%
4	Google News	1.49%
5	Google Maps	0.82%
6	Froogle	0.46%
7	Google Video Search	0.46%
8	Google Groups	0.43%
9	Google Scholar	0.27%
10	Google Book Search	0.25%
11	Google Earth	0.22%
12	Google Desktop Search	0.18%
13	Google Directory	0.10%
14	Google Answers	0.09%
15	Google AdWords	0.07%
16	Google Local	0.05%
17	Google Finance	0.03%
18	Google Calendar	0.01%
19	Google Talk	0.01%
20	Google Labs	0.01%

In May 2006, image search comprised almost 10% of Google search volume. Pair this with the knowledge that a smaller number of people on the Web optimize their sites properly for image search (or other vertical search engines) and you can see how paying attention to vertical search can pay tremendous dividends.

Of course, just getting the traffic is not enough. You also need to be able to use that traffic. If someone is coming to your site just to steal your image, for example, this traffic is likely not of value to you. So, although a lot of traffic may be available, you should not ignore the importance of determining how to engage users with your site. For instance, you could serve up some custom content for visitors from an image search engine to highlight other areas on your site that might be of interest, or embed logos/references into your images so that they carry branding value as they get “stolen” and republished on and off the Web.

Universal Search and Blended Search

In May 2007, Google announced Universal Search, which integrated vertical search results into main web results.

Thinking of it another way, Google’s web results used to be a kind of vertical search engine itself, one focused specifically on web pages (and not images, videos, news, blogs, etc.). With the advent of Universal Search, Google changed the web page search engine into a search engine for any type of online content. [Figure 8-1](#) shows some examples of Universal Search results, starting with a Google search on *iphone*.

Video results for iphone

 A Closer Look At The iPhone 4 min 25 sec www.youtube.com	 Will It Blend? - iPhone 1 min 37 sec www.youtube.com
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

News results for iphone

 Staying afloat in a sea of iPhone apps - 23 hours ago Just how lucrative is the iPhone App Store for a business in application d easy to achieve success or is money on app development better ... Computerworld - 651 related articles » Would Apple and AT&T Cripple a Google Voice iPhone App? - Wired News - 130 related articles » Operator O2 may lose exclusive distribution of the iPhone - Times Online - 326 related articles »

FIGURE 8-1. Search results for “iphone”

Notice the video results (labeled “Video results for iphone”) and the news results (labeled “News results for iphone”). This is vertical search being incorporated right into traditional web search results. [Figure 8-2](#) shows the results for a search on *i have a dream*.

Right there in the web search results, you can click on a video and watch the famous Martin Luther King, Jr., speech. You can see another example of an embedded video by searching on *one small step for man* as well.

The screenshot shows a Google search results page for the query "i have a dream". The search bar at the top contains the query. Below it, there's a "Web" filter button and a "Show options..." link. The first result is a link to "American Rhetoric: Martin Luther King, Jr. - I Have A Dream", described as "Full text and video and audio of Martin Luther King's I Have A Dream speech." The second result is another link to the same website, also titled "American Rhetoric: Martin Luther King, Jr. - I Have a Dream", described as "Text and audio of King's speech delivered at the Lincoln Memorial on August 28, 1963." The third result is a link to "The I Have a Dream Speech - The U.S. Constitution Online ...", dated Oct 2, 2008, described as "The text of the I Have a Dream Speech given by Martin Luther King in 1963." Below these results, there's a section titled "Video results for i have a dream" featuring three video thumbnails from YouTube and Google Video.

FIGURE 8-2. Search results for “i have a dream”

The other search engines (Yahoo!, Microsoft, and Ask) moved very quickly to follow suit. As a result, the industry uses the generic term *Blended Search* for this notion of including vertical search data in web results.

The Opportunity Unleashed

As we noted at the beginning of this chapter, the opportunity in vertical search was significant *before* the advent of Universal Search and Blended Search. However, that opportunity was not fully realized because many (in fact, most) users were not even aware of the vertical search properties. With the expansion of Blended Search, the opportunities for vertical search have soared.

However, the actual search volume for <http://images.google.com> has dropped a bit, as shown in Table 8-2, which lists data from Hitwise for February 2009.

TABLE 8-2. Most popular Google properties, February 2009

Rank	Name	Share
1	Google	68.42%
2	Gmail	9.62%
3	YouTube	9.41%
4	Google Images	5.76%
5	Google Maps	2.08%
6	Google News	1.38%
7	Google Video	0.54%
8	Blogger	0.46%
9	Google Calendar	0.43%
10	Google Groups	0.38%
11	Google Book Search	0.31%
12	Google Docs & Spreadsheets	0.26%
13	Google Finance	0.23%
14	Google Earth	0.20%
15	Orkut	0.09%
16	Google Scholar	0.08%
17	Google Pack	0.07%
18	Picasa by Google	0.04%
19	Google Answers	0.04%
20	Google Checkout	0.04%
21	Google Chrome	0.03%
22	Google AdWords	0.03%
23	Google Code	0.02%
24	SketchUp	0.02%
25	Google Desktop Search	0.02%
26	Google Directory	0.01%
27	Google Base	0.01%
28	Google Talk	0.01%
29	Google Groups 2 Beta	0.00%

Rank	Name	Share
30	Google Labs	0.00%
31	Google Web Accelerator	0.00%
32	Froogle	0.00%
33	Google Moon	0.00%
34	Google Catalogs	0.00%

This drop is most likely driven by the fact that image results get returned within regular web search, and savvy searchers are entering specific queries that append leading words such as *photos*, *images*, and *pictures* to their search phrases when that is what they want.

For site owners, this means new opportunities to gain visibility in the SERPs. By adding a blog, releasing online press releases to authoritative wire services, uploading video to sites such as [YouTube](#), and adding a [Google Local](#) listing, businesses increase the chances of having search result listings that may directly or indirectly drive traffic to their sites.

It also means site owners must think beyond the boundaries of their own websites. Many of these vertical opportunities come from one-time engagements or small additional efforts to maximize the potential of activities that are already being performed.

Optimizing for Local Search

Search engines have sought to increase their advertiser base by moving aggressively into providing directory information. Applications such as Google Maps, Yahoo! Local, and Bing Maps have introduced disruptive technology to local directory information by mashing up maps with directory listings, reviews/ratings, satellite images, and 3D modeling—all tied together with keyword search relevancy. This area of search is still in a lot of flux as evolutionary changes continue to come hard and fast. However, these innovations have excited users, and the mapping interfaces are growing in popularity as a result.

Despite rapid innovation in search engine technology, the local information market is still extremely fractured. There is no single dominant provider of local business information on the Internet. According to industry metrics, online users are typically going to multiple sources to locate, research, and select local businesses. Traditional search engines, local search engines, online Yellow Pages, newspaper websites, online classifieds, industry-specific “vertical” directories, and review sites are all sources of information for people trying to find businesses in their area.

This fractured nature of online local marketing creates considerable challenges for organizations, whether they’re a small mom and pop business with only a single location or a large chain store with outlets across the country.

Yet, success in these efforts is critical. The opportunity for local search is huge. More than any other form of vertical search, local search results have come to dominate their place in web search. For example, [Figure 8-3](#) shows the results for a search on *minneapolis rental cars*.

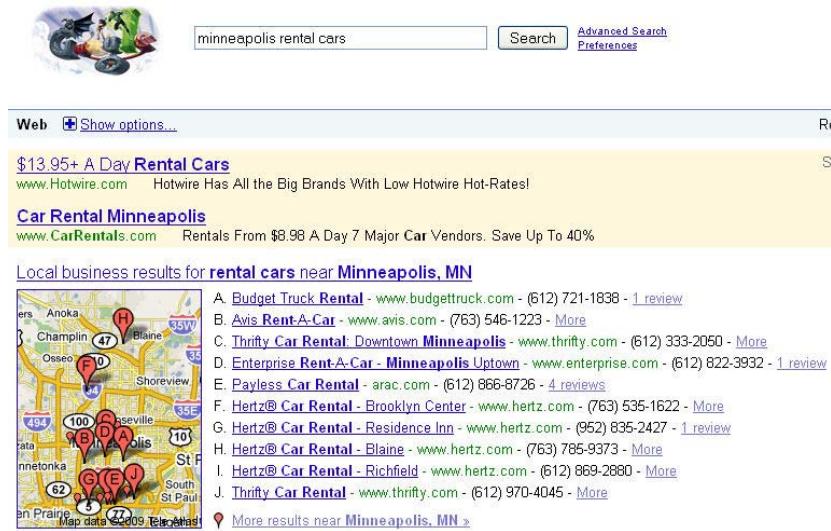


FIGURE 8-3. Local search results example

The regular web search results are not even above the fold. This means that if you are not in the local search database, you are probably not getting any traffic from searches similar to this one.

Obviously, the trick is to rank for relevant terms, as most of you probably don't offer rental cars in Minneapolis. But if your business has a local component to it, you need to play the local search game.

Foundation: Check Your Local Listings

Today, literally thousands of online directories and websites offer up guides to local businesses. So, if you have a local business or a chain of shops, where do you start?

Directories can be built from the local phone company's database information, but no one phone company covers the entire country. Because of this, companies that host nationwide directories are primarily getting their content from *data aggregators* to form the foundation of their guides. Data aggregators build their content from a variety of sources, such as local area Yellow Pages, to have information that is as comprehensive as possible.

Three top aggregators exist for U.S. business listings: [InfoUSA](#), [Acxiom](#), and [Amacai](#), as shown in [Figure 8-4](#).



FIGURE 8-4. Aggregators of business listings

The first step in managing the online presence of local companies is to check and update the business's listing information in each of these main aggregators. Ensure that the address information, phone numbers, fax numbers, and any other contact information are correct.

It is also a good idea to check/update your listing information in the top Yellow Pages directory sites, vertical directories (directories that are apropos for your industry), and top local search engines. But how do you decide what the top local information sites are?

As of this book's printing, the best guide for Internet Yellow Pages, vertical directories, and local search engines is the [Local Search Guide](#) provided by the [Yellow Pages Association](#). Check your listings in each of the sites listed in the guide, and update where necessary.

Of the search engines listed in the guide, you'll focus primarily on their local search or map search sections, and you'll want to look for how to add/update/edit your listings in them. For instance, you can find the interface for updating listings in the Google, Yahoo!, and Bing search engines at the following URLs:

- Google Local Business Center: <http://www.google.com/local/add>
- Yahoo! Local: <http://listings.local.yahoo.com/>
- Bing Local Listing Center: <https://ssl.bing.com/listings/ListingCenter.aspx>

Google and Microsoft offer methods for validation using automated phone and fax systems. Yahoo! offers a paid service in which the business pays for data maintenance and then gets direct control over its listings.

The major advantage that these systems offer is that you are validating your data directly with the search engines themselves. You can bet that they will treat this as highly trusted data. If the address you validate directly with them is 39 Temple Street and the InfoUSA address is 41 Temple Street, they are probably going to use the 39 Temple Street address.

Additional local info guides

Search engines are not the only source for local business information. Some of the more notable alternatives include the following.

Additional local online Yellow Pages. In addition to the online directories listed in the Local Search Guide, check to see that you've also updated your information in any local directory sites that are independent of the Local Search Guide lists. Other Yellow Pages guides may be dominant for your area but may not be listed. Check the printed phone books delivered in the area where your business is located, and see whether they have URLs printed on their covers where you can audit/update your information.

Additional vertical directory sites. The Local Search Guide lists only a handful of vertical directories, so if your industry isn't represented in that set, you might do some research to identify ones appropriate for you, and check them to ensure that your business listing is optimal in them.

Newspapers. Check the sites of the top newspapers in your area and see whether they have business directories with a good presence for you.

Chambers of commerce. Most U.S. cities have a local chamber of commerce to help promote businesses in the area, and getting listed within it can be beneficial to you, particularly if the chamber's site is optimized for search engines; getting your chamber of commerce listing linked over to your website can help with your link weight.

Online classifieds and eBay. These sorts of sites can be time-consuming to integrate with, but users sometimes conduct local-based searches through them for some types of products and services. [Craigslist](#) is the most-used online classifieds site, although there could be more specialized ones for particular cities or industries. [Figure 8-5](#) shows results from an eBay search for *stuffed animals*.

[EBay's](#) advanced search features allow users to search for things offered by sellers in particular regions/localities. So, for some types of businesses, it could be helpful or worthwhile to list products on eBay. Listing items through online classified or auction sites might not be good for improving direct sales, but it could be worthwhile as another channel for advertising to local consumers.

Local guides. Loads of local guides are devoted to information about local areas, so search on your city's name or zip code and see what sites appear on the first page of results in each of the main search engines: Google, Yahoo!, and Bing. Review the top local guide sites for your area and assess whether they're apropos for your business's information.

Specialty Yellow Pages. Many niche Yellow Pages directories are geared toward particular demographic groups—for instance, special interest groups or directories in other languages. Consider integrating with the ones that are right for you and your business. Association with these specialized guides may position you for more ready acceptance by the end users of those

	New 2 JUMBO TOY HAMMOCKs Net Organize Stuffed Animals	Buy It Now	\$19.95	Free shipping	20d 9h 31m
	Disney Animal Friends Stuffed Soccer Kids ball C0053	1 Bid	\$0.99	Free shipping	2h 17m
	NEW Disney Nightmare Before Christmas Stuffed Animal	3 Bids	\$5.77	Free shipping	2h 28m
	WEBKINZ LIL'KINZ GOOGLES STUFFED ANIMAL**UNUSED CODE**	Buy It Now	\$5.96	Free shipping	17d 8h 37m
	Care Bears stuffed animal game Skin for Nintendo DSi	0 Bids	\$9.92	Free shipping	2h 41m

FIGURE 8-5. Online classified sites

guides, because it sends them a clear message that you value their interests and are more directly sympathetic to their needs and desires. Here are some examples:

- Christian Yellow Pages
- Jewish Yellow Pages
- Black Business Planet
- National Black Yellow Pages
- National Green Pages
- Indian Yellow Pages
- Hispanic Yellow Pages
- Dog-friendly businesses

Introduction to Local Business Profiles

Many online directories and local search engines are adding more dimensions of information onto a business's basic listing. Providing as much detailed information about your company as possible through these profiles could be beneficial in terms of converting users of those sites into new customers for you. As you pass through each local search engine and directory to check and update your listings, notice what other opportunities may be available to you to enhance your business's information (see [Figure 8-6](#)).

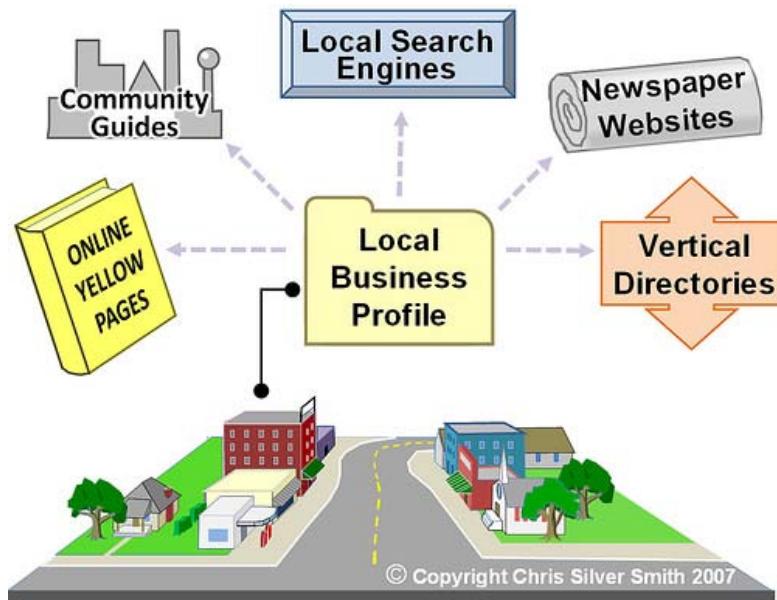


FIGURE 8-6. Role of local business profiles

Such enhanced profile information can include things like hours/days of operation, products and manufacturers carried, special offers, years in business, photos, video, slogans, business categories, keywords, certifications, menus, amenities, and payment methods, among others.

Add the URL pointing to your company's website to any/all of these local information websites. Nearly all local information websites NoFollow their links, so they do not have a direct SEO benefit, but you do want to make it easy for visitors to these websites to get to your site.

You'll notice that some directories and search engines are compiling all of this profile information from a variety of sources, so be aware that information supplied to one local information site could automatically begin appearing in many other places.

We will discuss the subject of optimizing your local business profiles in much more detail shortly.

Local Agency Management

Some companies may be overwhelmed by the prospect of updating and monitoring information through the many avenues we listed earlier. For a small business with just a few locations, the main work is in the initial integration and updates—thereafter you just need to periodically check and intervene as necessary. (You don't have to do your updating all at once. You could make a list of the most important online sites where your information may appear and methodically work through them as time allows.)

But checking and updating information for dozens to thousands of locations would not really be feasible to do by hand. In those cases, it would probably be sufficient to just focus on the top Yellow Pages, the business information aggregators, and the top search engines—from those information sources, your proper information would likely trickle down to all the rest. If your company has a substantial number of brick-and-mortar stores, you could contact each of the top Yellow Pages, aggregators, and search engines, and offer to supply them with a periodic feed of your business listings.

If you're still overwhelmed at the prospect of handling the update duties we previously outlined, a number of companies have sprung up to provide a management service for local profile information; if your business has many locations, you should review the cost/benefit of outsourcing the management of your data to one of these specialists. Here are the top three:

- [Localeze](#)
- [R.H. Donnelly](#)
- [GetListed](#), which has some additional DIY options

Optimizing Your Website for Local Search Engines

If you have been around for a while, your business probably is already included in the local search engines, since they compile data from the aggregators and other online directories. Once your business's listing is loaded into the local engines, you must determine how to get your business to rank higher when users search for your industry's keywords. On the next few pages, we will outline things you can do on your website to prepare it for better rankings in local search engines.

All of the basic SEO factors can come into play and can help to influence your rankings. These factors include having good, specific text in each page's title, H1 tags, meta description, page content, IMG ALT attributes, inbound links, and so forth. But some things are specific to local search, such as the following:

- If your company has multiple locations, it is not necessary to have a standalone website or subdomain (e.g., *loc1.example.com*, *loc2.example.com*, *loc3.example.com*, etc.) for each outlet. In fact, it is probably better if you don't, since each business location would likely produce similar website content. However, it probably would be helpful for you to create a separate profile web page on your site for each distinct location. Many sites with chain outlets will list all outlets on one page—that is not optimal. It is better to have one page be about one store location and another about a different location so that you can leverage all the on-page elements to create a page focused on that location.
- Have your page title, H1 tags, and content include the business name, type of the business, and location name—for example, “Acme Café: French Restaurant in Boston, MA.” For multiple locations, make the title different on each location's page. Include the neighborhood, street address, area nicknames, and other location-distinguishing information.

- The home page and/or Contact Us page should have the main location's basic listing information (street address, city, state, zip code, phone numbers, etc.) displayed somewhere in the HTML text. You should also add the basic listing information in the *hCard microformat*, which is a method for encoding address information on web pages for this purpose (learn more about it at <http://microformats.org/>). If you have multiple locations, display the basic information on each location's profile page.
- Place differentiating information on each store's pages, including a map, hours of operation, brand names, product photos, "in business since x" information, menus with prices (if it is a restaurant), ratings (if applicable), certifications, bonded status, and similar information. Of course, include specifics about the physical location, as we previously mentioned.
- Be aware of your proximity to your city's *centroid* (the location that the search engine defines as the center of the city). This is likely something that you cannot change at this point, but if you were considering moving your company to a new location, take into account where your city's centroid is located, and try to find a place closer to that point. Many of the map search engines will display the businesses located closest to the centroid first for any particular category or keyword search. Note that this is something that all the search engines now downplay as a factor, as it is less important than it used to be.
- Proximity works the same way for zip codes. If a user searches for businesses within a specific zip code, the businesses closest to the zip code area centroid will likely be displayed first in the list. Note, though, that zip code searches are not commonly performed in the United States (you can verify this with your favorite keyword research tool).
- It is very nearly a secret that good user ratings are one of the biggest factors for ranking high in a number of the local search engines, particularly Google Maps. Google Maps has compiled ratings from many other directory sources to get a universal rating for a business, so having high average ratings in Judy's Book, CitySearch, Superpages, [Yellowpages.com](#), and others will help improve your chances to rank higher in map search results.

Google has its own user ratings to combine into the mix as well. But how do you improve these ratings? As always, Google does not want you to artificially influence the reviews, and most of the online ratings have varying degrees of fraud detection, so Google would likely discount reviews if it believes they are being manipulated.

Do not attempt to set up multiple user accounts to rate yourself well or reduce your competition's ratings, though! Likewise, do not pay customers for beneficial reviews—if such a practice were discovered, you could lose all of your beneficial ratings. Focus on working in ways that are allowed under all the search engines' rules: just ask people to rate you. By asking enough happy clients to rate you, you might be able to drive up the positive reviews, which can bring rankings benefits. One thing you can do is post a sign on your premises encouraging customers to review your business on Yelp and other places, as well as post a Yelp widget on your site.

As part of good reputation management, you should monitor your online ratings in the various sites, and try to respond to complaints/issues as quickly as possible.

- Develop external links pointing to your website. The best links to support local search come from other locally and topically oriented sites with locally and topically oriented anchor text. For example, “Atlanta rental cars” is better anchor text than “rental cars.” Of course, all natural inbound links are beneficial.
- If your local business has its own blog, add a blog map or *feedmap* (<http://www.feedmap.net>) to it. This will add a local signal to the blog, as well as bring it to the attention of other bloggers in your area who are also participating in feedmaps.
- For businesses such as restaurants, it may be beneficial to add photos of celebrities who have visited the establishment.

This could translate well online, too—people perform searches on celebrity names all the time, so if you are lucky enough to have celebrities frequent your restaurant, uploading photos of them can help attract celebrity watchers to your business. The main thing is that it could drive up inbound links, and get users to linger longer on the site, perhaps bolstering the site’s quality scores.

- Community interaction can assist in online promotion. One great way to obtain local area links is to support charitable efforts in the area and to sponsor local events. Although it may seem crass to get link value out of charitable efforts, this is one of the best ways to support your local community while bolstering your business in the online marketplace. Don’t feel guilty for creating a win-win situation! Such opportunities could include:
 - Local charity races/walks, particularly if they list sponsors or sponsor logos on a page on their site. Request to have those linked to your local business site. Inbound links from these sites are very beneficial, and this is a valid way to get them!
 - Sponsorship of the local college or high school sports teams and bands. Also, request that they list sponsorships on the school website.
 - Local fraternal organizations, which may offer opportunities for links. Examples include the Freemasons, Lions Club, and Shriners.
 - Charitable events that you host yourself. Is anyone organizing the provision of food for the needy at Thanksgiving? If not, organize it yourself and host the information pages off your own website. This could draw many others to link to you as they help in promotion efforts.

Optimizing for Image Search

Optimizing for image search can be a smart strategy for many search marketers. Even if you’re working on a site that you don’t feel truly lends itself to an image optimization strategy, you may be able to leverage images or photos for SEO.

However, we should note that for some sites, there may not be a lot to gain here. Some e-tailers report poor conversion on image search traffic, and lots of the people coming in appear to be focused on stealing their images. You need to weigh the benefit of image search capability against the costs and other opportunities for SEO on your site. Nonetheless, many sites do very well with image search.

A significant amount of traffic can come from image search, and the number of people competing effectively for that traffic is much lower than it is in general web search. Industries that don't immediately seem to provide compelling subjects for images may enjoy greater potential in this area, because the competition might never clue in to the advantages of integrating images into their sites and into an overall search marketing strategy.

There are a few different ways that image search optimization can help to improve traffic and conversions for your site:

Subtle reputation management

Images of your products/services/facility assist consumers during the research phase of their shopping, and lend an implicit message of openness/forthrightness to your business. Providing generous numbers of images says you don't have anything to hide, and will improve consumer confidence in your company, increasing the chances that they'll select you to do business with.

Shopping via image search results

Increasingly, consumers are searching for products via image search engines because they can rapidly find what they are seeking without having to dig through promotion-laden websites. If your products can be found in the image search engine, you have an improved chance of being found by those people. With no pictures, there's zero chance of being found in image search.

Increase your chances of showing up more frequently in Universal Search

Performing image search optimization improves your chances of showing up in additional positions on the main search results pages as Universal Search pulls image search content into the main SERPs for some keyword search terms.

Empower others to promote you

If you have a flexible enough organization and you hold the legal copyrights to your images, you can allow others to take and reuse the images in return for promotion of your site/business.

Image Optimization Tips

Unlike normal web pages, which are often rich with text content, image search is much more difficult for search engines to perform. The image itself provides few clues to the content within it. Although search engines are experimenting with techniques such as optical character recognition (OCR) to read text content within images, most images don't have any text to read.

Search engines are also making use of facial recognition software to be able to determine when an image is of a face versus a body, or something else entirely.

However, although these types of technologies are very useful, they are limited in terms of what they can do. For that reason, success in image search optimization depends on using all the signals available to you to increase the search engines' confidence in the content of an image.

This certainly includes the basic SEO techniques we have discussed in this book. The web page's title tag, the H1 heading tag, the on-page content, and links to the page are all factors in image ranking. For example, if you have a single image on a page and it is a picture of the Golden Gate Bridge, and the title, heading tag, and page content all support that, the search engines' confidence in the content of the image increases. The same is true if you have 10 images on a page of many different bridges, and that is reinforced by the title, heading tag, and content. Consistency of content and theme is important in all of SEO, and it is especially critical in image SEO.

You should give particular emphasis to the text immediately preceding and following the image as well. This is what the user most closely associates with the image, and the search engine will view it the same way. A descriptive caption underneath the image is helpful.

You can do a number of things to further optimize your images. Here are the most important things you can do:

- Make sure the image filename or `IMG SRC` string contains your primary keyword. If it is a picture of Abe Lincoln, name the file `abe-lincoln.jpg` and/or have the `SRC` URL string contain it, as in `http://example.com/abe-lincoln/portrait.jpg`.
- Always use the image `alt` attribute. The `alt` attribute helps the vision-impaired to understand your site, and search engines use it to better understand what your images are about. Recent research by the authors indicates that this feature is still not used for lots of sites' images, and that many sites have tried to use it with invalid HTML. Make sure the `alt` parameter is valid, as in this example:

```

```

Use the quotes if you have spaces in the text string of the `alt` content! Sites that have invalid `IMG` tags frequently drop a few words without quotes into the `IMG` tag when they were intended for the `alt` content; with no quotes, all terms after the first word would be lost, if any would be used at all.

- Avoid query strings for `IMG SRCs` just as you should for page URLs. Or, if you must use URLs that include query strings, use only two or three parameters. Consider rewriting the query strings in the URLs so that they do not contain an excessive number of parameters, which will cause spiders to refuse to crawl the link. Note that Google claims to no longer have problems with these types of situations, but it is better to be safe than sorry, and other search engines have been less clear as to whether this is still an issue for them.

- Use good-quality pictures, which will read well when shown in thumbnail format. Good contrast is typically the key here. Lower-contrast images are visually harder to read, and it is common sense that if the thumbnail image doesn't look good, it will not invite a click.
- Do not save images as graphics files with embedded thumbnails—turn this feature off in Photoshop and other image editing software. Search engines may copy your image, reduce it in size, save it in compressed format, and deliver up a thumbnail of it for their results pages. An embedded thumbnail can wreak havoc with some compression software, and it increases your file size slightly, so just leave that feature turned off.
- Don't store the image in a sidebar column with your ads or inside the header/footer navigation elements; otherwise, the engine algorithms will ignore the image as irrelevant, just as they ignore page decor and navigation graphics.
- Have a proper copyright license! You need to have a proper license to display the images found on your site so that you don't get sued. Be careful about trying to use images from [Wikimedia Commons](#) or other public stock photo sites since you cannot be sure that those images really are in the public domain.
- If you are using images that may also be displayed on other websites, store/display them at different sizes from how they were provided to you. Don't change only their HTML `IMG` tag height/width parameters; reduce the size of the images or increase or decrease their compression and then resave them so that they have different file sizes.

Consider adding a watermark with your site URL to the image. This can bring traffic to your site, and it also discourages people from stealing your images.

Also, try altering the aspect ratio of the images a little bit—that is, make sure the height-to-width ratio of the images is different from images on other sites. You can do this by slicing a few pixels off the height or width of the graphic and resaving it. You need to change the file sizes and aspect ratios only if you are redisplaying image files that may also be found on other websites. This is particularly the case for licensed content, such as movie graphics, news article graphics, affiliated content, and manufacturer product images.

By changing the image files somewhat, you will help to ensure that the search engines perceive your content as being sufficiently original, instead of throwing it out of the SERPs after hitting a duplicate content filter.

- You need to ensure that your server configuration allows your site's images to be displayed when called from web pages on other domains. Some system administrators have disabled this to keep people from displaying their images on other sites, and this could cause problems if you want your images displayed in search engine image search results pages. Likewise, make sure that your `robots.txt` file does not block the crawlers from accessing your image file directories.
- If it is a fit for your business, specify that others are free to use your images for online display as long as they link back to your website from a credit line below the image or adjacent to the image where they display your copyright notice. Enabling others to use

your photos invites more promotional attention when people wish to write about you in blogs or in news articles.

Optimizing Through Flickr and Other Image Sharing Sites

Flickr is one of the strongest image sharing sites in terms of search optimization potential. No other image sharing site has the same level of domain authority, crawlability, keyword-focusing signals, and cross-referencing potential (the ability to link to your website from photo pages).

Note that Flickr does allow you to link to your site, but those links will be **NoFollowed**. The main benefit to be gained by optimizing in Flickr is that a well-optimized page can rank in the search engines for key terms, and provide you with more than one listing on those terms in the search results.

In other words, you can have your main website rank and get up to two potential spots, and then use Flickr to compete for a third spot in the top 10 results. That is prime real estate and well worth chasing. In some circumstances, the Flickr result may even outrank your main site, as you can see in [Figure 8-7](#).

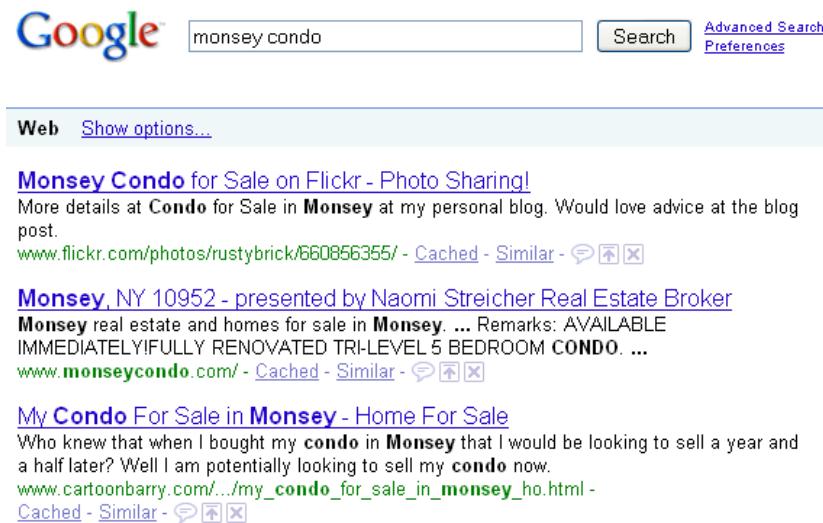


FIGURE 8-7. Flickr in Google's search results

In [Figure 8-7](#), originally posted by Barry Schwartz (<http://www.seroundtable.com/archives/014067.html>), the main site with the listing is www.cartoonbarry.com (the third result) and the photo of the condo on Flickr is the first result.

Other image sharing sites exist as well, even though they appear to have less potential. The following tips are specifically for Flickr, but most image sharing sites have similar features, so these tips could also work on many of them:

- When you upload your photos, always add tags. The tags or keywords that you associate with your photo will make the photo findable to users when they are searching and will lend keyword weight to the photo's page. Enter as many tags as possible that accurately should be keywords associated with your photo. However, make sure you place multiword tags within quotation marks (e.g., "pickup truck").
- The Flickr Tag Cloud, Flickr's user-tag "folksonomy," generates a good link navigation system for both users and search engine spiders.
- This should be obvious, but have your photos publicly viewable, not restricted to viewing by only your friends and family.
- Create a descriptive title for the image. This adds yet more keyword weight to the photo's page within Flickr.
- Enter a description under the photo, or write something about the picture.
- Consider adding a note or two directly onto the photo, particularly if the photo is humorous. Flickr allows you to select a rectangular area and associate some text with it that will appear as a tool tip when users mouse over it. Adding a humorous/interesting note or two may encourage users to participate on your photo's page, and the greater the participation/stickiness with the page, the better quality score the page may attain.
- If the photo is location-specific, geotag the picture.
- Create thematic sets for your photos, and add each picture to the set(s) appropriate for it. This provides yet more contextual clues for search engines regarding the content of the photo's page, and it will allow a user arriving at the page a way to easily look at similar pictures you've taken.
- Browse through Flickr's Groups for public "photo album" collections that are dedicated to pictures that could be related to your photo. Sometimes it helps to search for photos using keywords you used in your tags, and then see what groups other people's photos are members of.
- Join those groups, and then add your photos apropos to each group's theme. The more links to your photo page, the more important your photo will be considered, so add it to a number of groups. Ideally, add it to groups that have a lot of members—the number of members indicates the popularity and traffic of the group.
- Link each of your Flickr photo pages to your website or to related pages on your website. You can add hyperlinks to the Description field below the photo. Use anchor text that has a call to action, or that tells the user what to expect if he clicks on the link (e.g., "We sell this product on our website"; "Enjoy this view from the tables at our restaurant"; "This room is available at our bed & breakfast"). It is best to link to specific pages of related content as a richer indicator for link juice transfer.

- Finally, post as many optimized pictures as possible. This is mostly a game of many small fractions adding up to large, cumulative results. The many pages of pictures linking back to your site will help build your overall authority. The more pages you have, the more likely it is that other Flickr users will find your content and link to it. It also increases your chances that a lucky picture or two might find its way onto a viral popularity wave that spurs many users to send links of your picture to their friends, or that a reporter might find one of your pictures ideal for his news story.

Optimizing for Product Search

Google Product Search is less popular than some of the other Google-owned properties (reference the Hitwise charts shown at the beginning of this chapter). However, top rankings for product-based businesses in Google Product Search are critical. This is because the top three Google Product Search results sometimes make their way into the main web search results, grouped together as a *onebox* (specially grouped vertical search results added to the SERPs along with the traditional 10 blue links). This Product Search onebox can appear anywhere on the page—top, middle, or bottom. [Figure 8-8](#) shows an example for the search *canon powershot*.

canon powershot Advanced Search Preferences

Web Show options...

[Canon Digital Cameras](#)
Shoot with ease and style with the **Canon PowerShot** cameras including Digital ELPH series. **PowerShot** digital cameras incorporate the creative performance of ...
www.usa.canon.com/consumer/controller?act...113 - Cached - Similar

[PowerShot SD850 IS Digital Camera](#)
Canon, DIGIC, PIXMA and **PowerShot** are registered trademarks of **Canon** Inc. in the United States and may also be registered trademarks or trademarks in other ...
www.usa.canon.com/consumer/controller?act...145 - Cached - Similar
[More results from www.usa.canon.com >](#)

[Amazon.com: Canon PowerShot S400 4MP Digital Camera w/ 3x Optical ...](#)
Canon PowerShot SD1200IS 10 MP Digital Camera with 3x Optical Image ... The 4 Megapixel **Canon PowerShot** S400 digital camera is capable of taking still ...
www.amazon.com/Canon-PowerShot-S400.../B00008OE6I - Cached - Similar

[Amazon.com: Canon PowerShot SD850 IS 8.0 MP Digital Elph Camera ...](#)
Amazon.com: **Canon PowerShot SD850 IS** 8.0 MP Digital Elph Camera with 4x Optical Image Stabilized Zoom: Camera & Photo
www.amazon.com/Canon-PowerShot-Digital.../B000Q3D420 - Cached - Similar
[More results from www.amazon.com >](#)

Shopping results for canon powershot
[Canon PowerShot SD1100 IS Digital ELPH Digital camera - compact - 8.0 ... \\$164 new, \\$140 used - 76 stores](#)
[Canon PowerShot A590 IS Digital camera - compact - 8.0 Megapixel - 4 x ... \\$122 new, \\$76 used - 61 stores](#)
[Canon PowerShot SX10 IS Digital camera - compact - 10.0 Megapixel - 20 ... \\$345 new, \\$300 used - 173 stores](#)

FIGURE 8-8. Product search onebox for “canon powershot”

Getting into Google Product Search

The first step toward optimizing your website for Google Product Search is to put together a feed for your products and submit them to [Google Base](#). You can upload products in bulk, and you can learn the specifics of the formatting of the feed at <http://www.google.com/base/bulkuploads>.

To be included in Google Product Search you need to upload true physical/tangible products. Google Base will accept other types of items, such as flights, hotels, car rentals, travel packages, and real estate, but such items will not get into Google Product Search.

In your feed, populate as many fields as possible with data (fields such as Brand, Category, Color, Price, Condition, and more). These additional fields will help Google Product Search match you up with more potential customers. It is also important to update your feed as often as possible. Some major e-tail sites update their feeds on a daily basis.

Product search optimization

Here are some of the basic things you can do to optimize your feed for Google Product Search:

- Create descriptive, accurate item titles.
- Use long tail terms in titles, particularly for highly competitive products. Picking the right terms can be difficult at times because there may be hundreds of long tail terms to choose from. Try to pick terms that are a bit more likely to have some search volume, and that have a potential for conversion. You can use the long tail keyword research techniques we laid out in [Chapter 5](#) to find the best candidates. On large e-tail sites (with thousands of products) where hand-researching each page is not possible, pick out unique attributes of the products that users may choose to type in when searching.
- Seller ratings play a big role in rankings in Google. Manage your ratings at contributor sources such as DealTime, NexTag, PriceGrabber, ResellerRatings, and Shopzilla.
- Product ratings are also important. Get your products rated on sites such as [Epinions.com](#).
- It appears that product names plus brand names in item titles are the best choice. You cannot really invoke non-brand-name searches effectively in Google Product Search.
- Always include product images. Google Product Search has a preference for products that display a product image (in Bing, if you have no photo, your product simply will not be shown). Here are some tips for product images:
 - Google converts images to 90×90 pixels to display thumbnails. Therefore, square pictures will take better advantage of the space.
 - Ensure that the product is sized as large as possible in the picture.
 - Higher-contrast pictures are easier to read in thumbnail size. Make sure pictures are not muddy/low-contrast.
 - Make products appear clearly against backgrounds.

- Other factors that also may play a part include the following:
 - The perceived authority of the domain
 - Specific pricing details (may be used to sort results)
 - Website's ranking for the keyword in web search
 - Whether any of the products are deemed "adult" (if so, all of your products might get filtered out due to SafeSearch)
 - Users specifying Google Checkout items only
 - Number of users who have added your Google Product items to their individual Shopping Lists within Product Search, or placed them on their Shared Wish List

Performance reporting

As of July 2008, Google offers a Performance tab with reporting on how many of your items received clicks, as well as how many items were uploaded and how many are active. You can download the data as a CSV file. **Figure 8-9** shows the type of data you can get.



FIGURE 8-9. Google Product Search performance reporting

You can read more about performance reporting at <http://googlebase.blogspot.com/2008/07/get-feedback-on-your-items-performance.html>.

Optimizing for News, Blog, and Feed Search

News, blog, and feed search is another large potential area of opportunity for optimization. This has a bearing on obtaining traffic directly from the search engines, but also in promoting your business in whole new ways.

Getting plugged into news search, for example, can bring you plenty of quality traffic, but it can also result in your site being spotted by major media editors and writers, as well as bloggers who consume that media. This is the type of exposure that can lead to numerous links.

Blogs and RSS feeds offer a similar dynamic of getting your content in front of new readers through new channels. There is also a social aspect to blogging, due to the built-in mechanism for comments and the tendency for bloggers to interact heavily with each other.

Optimization for news, blogs, and feeds is an area that applies to a wide range of sites. First up is a look at the optimization of RSS feeds.

RSS Feed Optimization

Many people mistakenly lump blogs and RSS together, but RSS has infinitely more applications beyond just blogs! Examples include news alerts, latest specials, clearance items, upcoming events, new stock arrivals, new articles, new tools and resources, search results, a book's revision history, top 10 best sellers (as Amazon.com does in many of its product categories), project management activities, forum/listserv posts, and recently added downloads. A good place to start is with basic SEO practices, as we've outlined elsewhere in the book (e.g., good titles, good descriptions, handle tracking URLs properly, etc.).

Here are the basics for RSS feed optimization:

- If practical, use the full text of your articles in your feeds, not summaries. A lot of users want to read the full article in the feed without having to click through to your site. This is a case where you need to focus more on the relationship with the user than on immediate financial goals.
- Consider multiple feeds. You can have them by category, latest comments, comments by post, and so on.
- An RSS feed that contains enclosures (i.e., podcasts) can get into additional RSS directories and engines, as there are many specialized directories just for podcasts or other types of media.
- Make it easy to subscribe. Ideally, users should have to click only once to subscribe via their favorite aggregator. You can do this through "Add to ____" (My Yahoo!, Bloglines, Google Reader, etc.) buttons on your site. Also make sure to implement `<link>` tags for auto-discovery (e.g., `<link rel="alternate" type="application/rss+xml" title="RSS" href="http://feeds.stephanspencer.com/scatterings" />`) in the `<head>` section of your web pages.

RSS Feed Tracking and Measurement

There are some important tracking and measurement issues to consider when implementing RSS:

- You should be tracking reads by embedding a uniquely named 1-pixel GIF within the <content:encoded> container. This is known as a *web bug*. Email marketers have been using web bugs to track open rates for ages. An example web bug might look like this:

```

```

- You should be tracking click-throughs by replacing all URLs in the <link> containers with click-tracked URLs. You can code this in-house or you can use a hosted ASP service such as SimpleFeed to do this for you. Incidentally, FeedBurner offers imprecise counts based on users' IPs, not on click-tracked URLs. An example click-tracked URL might look like this:

<http://www.yourdomain.com/rss.php?88288Cafaa903cf09ecafb231662ca053c4c56>

One often overlooked area of RSS click tracking is how to pass on the link juice from the syndicating sites to your destination site. Use click-tracked URLs with query string parameters kept to a minimum, and then use a 301 redirect to send the user to the URL without the query string. Surprisingly, FeedBurner and SimpleFeed use 302 redirects by default, so if you use one of these, make sure you configure it to use a 301 instead.

- You should be tracking your circulation (number of subscribers). Again, you could use a service such as FeedBurner, which categorizes visiting user agents into bots, browsers, aggregators, and clients. Bots and browsers don't generally "count" as subscribers, whereas a single hit from an aggregator may represent a number of subscribers. This number is usually revealed within the user agent in the server logs; for example, Bloglines/2.0 (...; xx subscribers). Today, tracking subscribers is an inexact science.

Other RSS Optimization Considerations

A few additional factors play a role in optimizing your RSS feeds. Here are three of the most important ones:

- Sites using your feeds for themed content to add to their site for SEO purposes could strip out your links or cut off the flow of the link juice using the `NoFollow` `rel` attribute, or by removing the HREFs altogether. Scan for that and then cut off any offenders by blocking their IP addresses from accessing your feeds.
- You should own your feed URL (unless you want to be forever tied to FeedBurner or whatever RSS hosting service you are using). Remember the days long ago when people put their [Earthlink.net](#) email addresses on their business cards? Don't repeat that mistake with RSS feeds.

- Use [Pingomatic.com](#) to ping the major feed services, as it will notify all of those services every time you make an update to your blog. However, you can choose to go directly to certain major publishers, such as Yahoo!. If you do prefer to go directly to Yahoo!, use this URL: <http://api.my.yahoo.com/rss/ping?u=http://www.yourdomain.com/rss> (replace “yourdomain.com/rss” with the address of your feed). You can get more information on this from Yahoo! at <http://publisher.yahoo.com/submit.php>.

Something to keep in mind for the future of RSS is that there may be a strong shift toward personalized feeds. Having only one generic RSS feed per site is a one-size-fits-all approach that cannot scale. On the other hand, having too many feeds to choose from on a site can overwhelm the user. So, you may want to consider offering a single RSS feed, but one in which the content is personalized to the interests of the individual subscriber.

If the feed is being syndicated onto public websites, you will want to discover that (by checking the referrers in your server logs), and then make sure the RSS feed content is consistent from one syndicated site to the next so that these sites reinforce the link juice of the same pages with similar anchor text. Or simply ask the subscriber his intentions (personal reading or syndication on a public website) as part of the personalization/subscription sign-up process.

Personalized feeds are not ideal from an SEO standpoint because you are not reinforcing the same items across multiple sites (when that feed is syndicated and displayed). You can mitigate this somewhat by offering standardized feeds for your affiliates and other webmasters. You can also make a special affiliate feed that already has that affiliate’s tracking code appended to all the links. Then you can 301-redirect those URLs to the canonical URLs. Ultimately, the value of improving the user experience exceeds the loss of SEO benefits in this aspect of feed optimization.

Blog Optimization

Blogs are great publishing platforms for those who want to write articles on a regular basis. First, they make it easy to publish the content. Authors only need to log in and use a relatively simple set of menu choices to input what they want to publish, preview it, and then proceed to publish it. It is far easier than coding your own HTML pages by hand.

In fact, it is so easy that websites have been built using WordPress as the sole publishing platform for the site. They are also typically easy to set up and configure. The world’s most popular blog platform is [WordPress](#), but the blog platforms from [Moveable Type](#) and [TypePad](#) are also popular.

A host of social marketing benefits come from blogs. Blogs are inherently social in nature. Enabling comments allows for interaction with your readers, and bloggers tend to have a significant level of interaction. For example, one blogger may write a post that reacts to or comments on another blogger’s post. A lot of cross-linking takes place, with one blogger citing another.

Working this aspect of blogging as a social media platform is beyond the scope of this book. Nonetheless, be aware that a blog is an opportunity to establish yourself as an expert in a topic area, and to engage in a give-and-take activity that can dramatically change the visibility of your business.

In addition to these huge benefits, blogs can also bring you search engine and/or blog search engine traffic when they are properly optimized.

Structural blog optimizations

As we have discussed throughout this book, there are many key elements to successful SEO. These are things such as title tags, heading tags, good content, inbound links, and SEO-friendly architecture. Although the various blog publishing platforms are great, they can sometimes also require some tweaking to achieve optimal SEO results:

- Blogs usually offer the ability to categorize each post. Make sure the tag name is used in the title of that tag page.
- Override default title tags with custom ones. You can do this using plug-ins such as [Netconcepts' SEO Title Tag](#). This plug-in allows you to override the title tag with a custom one (defined through a custom field in a post or a page).
- Rewrite your URL to contain keywords, and to use hyphens (preferred over underscores) as word separators. Do not let the blog platform include the date in the URL. If you are concerned that too many words in the URL can look spammy, consider post slug shortening with a WordPress plug-in such as [WordPress Slug Trimmer](#) or [Automated SEO Friendly URL](#).
- Make sure you 301-redirect from <http://yourblog.com> to <http://www.yourblog.com> (or vice versa). Note that if you have a site at <http://www.yourdomain.com> and a blog at <http://www.yourdomain.com/blog>, you may need to implement a separate redirect just for the blog. This has to be handled not just for the home page, but also for all internal pages (e.g., Permalink pages). Each URL must redirect to the corresponding URL on the www version.
- If you change from one blog platform to another one, the URL structure of your blog will likely change. If so, make sure you maintain legacy URLs by 301-redirecting from each page location to the new one.

In general, be on the lookout for situations where your blog interferes with basic SEO principles. Working around these limitations can be the difference between mediocre and great results. [Figure 8-10](#) depicts a sample search result that shows the impact of using keywords in blog page titles.

Notice how the first four results have the keywords in their title as well as in the URL.



Web [Show options...](#)

[RSS Can Feed Your SEO Efforts - Search Engine Guide Blog](#)

RSS is one of the hottest phenomenons in the webmaster community today. RSS that benefits SEO is a topic that's been very much overlooked.

www.searchengineguide.com/todd.../rss-can-feed-your-seo-efforts.php - [Similar](#)

[RSS and SEO: Implications for Search Marketers](#)

Also at Search Engine Strategies, Stephan Spencer provides a list of things to consider when dealing with RSS, entitled "RSS and SEO: Implications for ..."

www.stephanspencer.com/.../rss-and-seo-implications-for-search-marketers -

[Cached](#) - [Similar](#)

[Web Content Syndication : RSS SEO Strategy : PRESSfeed](#)

Web content syndication RSS and SEO: RSS puts search engine optimization on steroids so you can dominate the search engine results for your keywords.

www.press-feed.com/howitworks/syndicatecontent.php - [Cached](#) - [Similar](#)

FIGURE 8-10. Impact of using keywords in blog page titles

Optimizing your anchor text

Anchor text is just as important in blogging as it is in general SEO. You need to leverage it as much as you can. Here are some specifics:

- Make the post's title a link to the Permalink page. You do not want your only link to the post saying "Permalink".
- Use a tool such as [Link Diagnosis](#) or [Linkscape](#) to see who is linking to your site. Using these tools or tools like them, you can see who is linking to you and what anchor text they have used. Look for opportunities to request revisions to anchor text on inbound links. Make sure you are comfortable that your relationship with them will not result in their simply removing your link instead of changing it.
- Internally link back to old, relevant posts within the body of a blog post. Don't use *here* or *previously* or similar words as the anchor text; use something keyword-rich instead.

Sticky posts

Sticky posts are a way to add content that always shows up first on a page of your blog. One way to use this is to write a post that is the introduction/overview for all the content on one of your category pages. Using this technique, you can add keyword-rich introductory copy to a category page or tag page.

Figure 8-11 is an example of a sticky post associated with a category page on a blog.



FIGURE 8-11. Example of a sticky post

Author profile pages

If you have a multiauthor blog, another smart tactic is to create author profile pages. This has a multitude of benefits. First, many of your readers will have a stronger interest in one of your writers than the others. This allows them to follow that writer more closely. Better still, offer RSS feeds on a per-author basis. In addition, authors are likely to link back to their author pages on your blog from their own sites. Figure 8-12 is an example of an author profile page.

More blog optimization basics

As we previously emphasized, a lot of basic SEO techniques also apply in optimizing your blog. For example, the use of emphasis tags within posts (bold, strong, em, etc.) can have a significant impact. In addition, use the `Rel=Nofollow` attribute on links where appropriate—for example, in all links in trackbacks, in comments, and in posts where you do not vouch for the site.

Links remain critical

Obtaining links and managing your link juice remain critical activities. Blog platforms provide limited ability to manage your internal link juice, so this may require some customization to accomplish. Fortunately, in the WordPress environment some really good plug-ins are available to help you with this.

The screenshot shows a website for "Business Blog Consulting". At the top, there are four service boxes: "The Computer Consultants", "Manage Your Team Online", "Great Plains", and "Small Business Consulting". Below these is a search bar and a "Home : Email Us" link. The main content area features a sidebar on the left with sections for "About This Site:", "Looking for a company?", "Other stuff?", and "Site Contributors:" (listing Erin Blaskie, Toby Bloomberg, and Andy Bourland). The main content area is titled "About Contributor Dana VanDen Heuvel" and contains information about Dana's posts, website, email, jabber, profile, and a detailed biography. To the right of the main content is a "Syndicate" section with an RSS feed icon, a reader count of 21,63, and a "Posts via e-mail" form. Below that is a "Recent Posts" section with links to various blog posts.

FIGURE 8-12. Example of an author profile page

For starters, you can implement tag clouds and tag pages. Tag clouds are designed to be navigation aids. The concept is that if you arrive at a site and it has tons of content and you just want to see what is most popular, you can do that quickly with a tag cloud. Figure 8-13 is an example of the tag cloud at Stephan Spencer's blog.



FIGURE 8-13. Example tag cloud

You can also create cross-links between related posts using a plug-in such as [Yet Another Related Posts Plugin](#). This is a great way to get people who just finished reading one post on your blog to consider reading another one.

Other things that are interesting to promote in one fashion or another are your top 10 posts. However, make sure each post links to the next and to previous posts using keyword-rich anchor text. [Figure 8-14](#) shows a blog that links to other related posts made in the past.

He admits these are his personal views and is not an indicator of things to come. But it is a good reminder to be ready for it if it materialises in future.

[\(more...\)](#)

 [ShareThis](#)

Possible Related Posts

Related posts:

1. [Quova Awarded Patent for Improved Geotargeting](#)
2. [The Future Of SEO](#)
3. [Social Media Marketing Goals](#)
4. [Google Image Search - Second Only To Web Search In Size](#)
5. [Measuring Link Bait](#)
6. [Domain Diversity - Key Metric For Higher Google Rankings](#)
7. [Search Engine Crawling and Indexing Factors](#)
8. [Natural Search Marketing Is Just in Time Conversation](#)
9. [SEO Best Practices - Correlation Testing](#)
10. [Is Verizon Responsible for iearc's Bankruptcy?](#)

Posted by [Ravi of Netconcepts Ltd.](#) on 07/26/2009

[Permalink](#) |  [Email This Post](#) |  [Print](#) | [Trackback](#) | [Comments \(0\)](#) | [Comments RSS](#)

Filed under: [SEO](#), [Search Engine Optimization accessibility of seo](#), [Add new tag](#), [auckland search marketing](#), [content seo](#), [correlation analysis](#), [data driven seo](#), [future of seo](#), [link graph analysis](#), [Netconcepts](#), [organic seo](#), [ppc services](#), [query demand](#), [social graph metrics](#), [social media optimisation](#), [usage metrics](#), [verticalization of seo](#)

Social Media Marketing Goals

As more and more companies have adopted SEO practices in optimizing their websites, though there is still plenty of room for improvement, it's no surprise that they have begun turning more attention to social media marketing (SMM). Of course, with all the attention these new venues have gained and become more mainstream, this only makes sense.

2009			
jan	feb	apr	may
2008			
jan	feb	mar	apr
may	jun	jul	aug
sep	oct	dec	
2007			
jan	feb	mar	apr
may	jun	jul	aug
sep	oct	nov	dec
2006			
mar	apr	may	jun
jul	aug	sep	oct
nov	dec		
2005			
jan	feb	mar	dec
2004			
may	jun	jul	aug
sep	oct	nov	dec

Other

- [login](#)
- [register](#)
- [subscribe](#)
-  [Blogs that link here](#)
-  [Technorati](#)

Recent Readers



You!
Join Our Community



100% Free

Marketing Consulting For Your Website

FIGURE 8-14. Links to related posts from the past

In-bound links remain critical in the world of blogs too. [Chapter 7](#) covers this topic in great detail in “[Social Networking for Links](#)” on page 321, but here are a few extra tips about link building in the world of blogging:

- Develop relationships with other bloggers, and get them interested enough in your content that they follow your blog and perhaps even add you to their blog rolls.

362 CHAPTER EIGHT

- Be aware that trackbacks and comments usually will not provide links that pass link juice, so don't spend your time planning to leverage these things for the purposes of link building (you may choose to use them for other purposes).

Can you do this?

Although all of this may sound difficult, it isn't necessarily so. Co-author Stephan Spencer's daughter launched a blog at the age of 14 and achieved good results (see Figures 8-15 and 8-16).

This blog could be run by you! August 27, 2008

Posted by Chloe in : Site News , 49 comments

Would you consider yourself a "neo-expert"?

Want to work for NeopetsFanatic.com? Become the main blogger, publishing posts on this homepage just a couple times a week, and get paid!

Do you love Neopets.com? Do you log on almost every day, and do everything you can to earn neopoints, like using the stock market, etc? Are you a "neo-millionaire", or have at least a couple hundred thousand neopoints? Do you know almost everything there is to know about Neopets and it's new re-do?

Then you definitely qualify!

Email Chloe with a paragraph or two on how you would be the perfect blogger for NeopetsFanatic.

FIGURE 8-15. Creating a great blog; it's not that hard

News Search Optimization

Most people are so conditioned to the fact that Google is the dominant provider of search that few of them realize that the king of news search engines is Yahoo! News. Figure 8-17 shows data published by Hitwise that indicates the market share of various news portals.

As you can see, in a complete role reversal, Yahoo! News gets nearly three times as many visitors as Google News does. If you are able to implement a true news feed on your site (which requires a substantial commitment of time and resources), you should submit to both Yahoo! News and Google News. Here are their submission URLs:

- Yahoo! News: <http://help.yahoo.com/l/us/yahoo/news/forms/submitsource.html>
- Google News: www.google.com/support/news_pub/bin/request.py

The screenshot shows a search results page from a web browser. At the top, there is a logo featuring a dragon and a search bar containing the query "neopets cheats". Below the search bar are buttons for "Search" and "Advanced Search Preferences". The main content area has a light blue header bar with the text "Web" and a "Show options..." link. The first result is titled "Neopets Cheats" and describes it as a site for Neopets games. It includes links to "www.neopetscheats.org/" and "Cached - Similar". The second result is titled "Neopets Cheats : Cheat Codes and Walkthrough Guides for Neopets" and describes it as a site for Neopets games. It includes links to "www.neopets-cheats.net/" and "Cached - Similar". The third result is titled "Neopets Cheats, Game Help, Neopets Freebies, Neopets Dailies | The ..." and describes it as a Neopets help site. It includes links to "www.thedailyneopets.com/" and "Cached - Similar". Below these results, there is a section titled "Video results for neopets cheats" with two video thumbnails. The first video is titled "neopets cheat" and is 1 min 57 sec long, from www.youtube.com. The second video is titled "Neopets cheats" and is 3 min 55 sec long, from www.youtube.com. At the bottom of the page, there is a link to "The Ultimate Neopets Cheats Site" which is described as a game cheats, freebies, shop backgrounds, guild layouts, tips, and a weblog maintained by a Neopets addict. The link "www.neopetsfanatic.com/" is circled in red.

FIGURE 8-16. The Neopets Cheats site from Figure 8-15, ranking #5 for its trophy term



FIGURE 8-17. News search market share (Hitwise)

In doing so, you can gain a lot in terms of relevant traffic. In May 2007, Loren Baker of Search Engine Journal reported obtaining up to 5,000 visitors per day from Yahoo! News alone (<http://www.searchenginejournal.com/how-to-submit-your-site-to-yahoo-news/4971/>). That is a serious bump in traffic.

Of course, the exact numbers will vary from one situation to another, and even in Loren's example the results changed from day to day based on the degree of placement provided by Yahoo! and the keywords involved.

Note, though, that the timing of Loren's report was prior to the search engines' roll-out of Blended Search. With the integration of news search results into the engines, this opportunity has likely increased substantially. Suffice it to say that if you can generate the right type of content, there are plenty of reasons to optimize it, and to submit it to the news search engines.

Optimizing for news search

The news search engines are looking for content that is in the form of either a news story or a feature story. They also are looking to see that you are creating news content in reasonable volume—a minimum of 10 articles per week. News sites are looking for news sources (i.e., sites), not individual news pieces.

Guidelines for the content are the same as they are for traditional news. The articles should have a catchy, keyword-rich headline and a strong opening paragraph.

The opening paragraph should draw the reader in so that he will read the rest of the article. In traditional news, the main compelling point is put forth at the start, and the discussion continues through other points of descending importance. The news piece should end with a strong concluding paragraph that reviews the major points of the article.

Submission details

Google News has some more specific requirements than Yahoo! News, so a best practice is to make sure you meet Google's minimum bar. Here are some of the things you will need to get into Google News:

- All news stories should appear at a static URL. The URL and the content on it should not change over time. The URL also needs to be accessible through a standard HTML text link.
- Create a [Google News Sitemap](#) and submit it to Google through Google Webmaster Tools.
- Keep all content in standard HTML (no PDFs, multimedia, or frames). Make sure the articles are UTF-8-encoded.

When dealing with the form itself, here are some things you should plan to provide:

- A history of the site, including traffic statistics. Sell a little bit, so if your site has won awards or gotten recognition of any kind, include that information. This includes mentioning major sites that link to you.

- A two- to three-sentence background on your editors and authors. You will need at least three of these to be considered. You can provide this on your news page or link to a page with this information from your news page.

Google News Sitemaps can speed up the discovery process for your news articles as well. You can implement this as an RSS or Atom feed and have it ping the news server. Even without this, it is still helpful to create a News Sitemap file for your news. The format of this feed differs from the traditional XML Sitemaps file, as detailed in the link we provided earlier in this section.

Yahoo! News prefers to work with an RSS feed, so make sure you have such a feed ready when you submit to Yahoo!. Although Google News has more submission requirements, Yahoo! News can actually be more difficult to get into. Google News will accept a wider range of decent-quality news sources. Yahoo! News will not add your feed unless it feels you are bringing significant quantities of high-value content to the table that is not available from other sources.

Sitemaps and RSS feeds

Yahoo! News accepts RSS feeds, which should rapidly increase Yahoo!'s awareness of any new story. It is an excellent idea to make use of this. Standard RSS feeds should work.

Others: Mobile, Video/Multimedia Search

Mobile and video search are two areas experiencing rapid growth. Each has its own challenges and requires a unique approach.

Mobile Search

Mobile search optimization is a growing market. Outside the United States, where mobile device usage has a much higher level of penetration, this discipline is already very important. Within the United States, mobile device usage is growing rapidly, so it is in the process of becoming very important there as well. [Figure 8-18](#) shows an example of a search result on an iPhone.

Improved mobile-web-user interfaces on mobile devices are a big factor in the growth of mobile search. iPhone users conduct an average of 50 times more searches on Google than on Google's nearest competitor (http://blogs.computerworld.com/iphone_users_search_google_5000). The disproportionate usage of Google from iPhones is in no small part due to the exceptional ease of use and easily accessible Google search, and with the Google Mobile app, you can even search with your voice.

Google Mobile employs a different spider than Googlebot does, with the user agent string "Googlebot-Mobile". Google Mobile's ranking algorithms are separate as well. When Googlebot-Mobile crawls your site, it will let your web server know that it prefers to receive the mobile version of that page by sending an HTTP Accept request header. If your web server is properly



FIGURE 8-18. Search: different on mobile devices

configured for content negotiation and you have a mobile version of the requested content, Googlebot-Mobile and mobile users alike will receive your mobile content by default without having to be redirected to a mobile-specific URL.

As a result, you may wish to offer a mobile-specific URL for marketing and promotion purposes or to allow access to your mobile version regardless of the user agent making the request. Some of the more well-known examples of such URLs include <http://m.flickr.com>, <http://m.facebook.com>, <http://mobile.yelp.com>, <http://mobile.weather.gov>, <http://www.hotels.com/iphone>, and <http://en.m.wikipedia.org>. The benefit of serving up a distinct mobile version instead of serving your standard version with a mobile-friendly style sheet (CSS) is that you can trim the file size of the HTML sent to the mobile device.

Here are some important ranking factors that are specific to mobile SEO:

- Small, lightweight, and fast-loading site (<20 KB per page)
- XHTML Mobile 1.0 DOCTYPE
- UTF-8 character encoding
- JPEG/GIF images
- On-site keyphrase optimization as usual (with a focus on short titles and small amounts of body copy); include the word *mobile* in the title or the heading tag (or both)
- Use of Mobile Sitemaps

Other than these factors, it is important to follow the normal rules for SEO, bearing in mind that the page size is inherently smaller because of the size of the displays on the mobile devices. Although mobile-device experiences are likely to continue evolving so that the differences between a mobile browser and a standard web browser will become less significant, the issues with a smaller form factor are not likely to go away entirely.

Video Search Optimization

Video search optimization is in its infancy, but it is growing rapidly. The increasing penetration of high-bandwidth Internet connections is going to drive more and more demand for online video content. So, creating video content and optimizing it is something you should consider.

The problem is that video search engines have not yet been able to look inside the video to tell what the content is from a human “eye” perspective. For that reason, they rely on other on-page factors, the metadata within the video file itself, and inbound links and anchor text to determine what the video is about.

N O T E

There are exceptions to the statement that video search engines cannot look inside the video. They can do some basic things such as flesh detection to identify adult content, face detection, on-screen text detection, and so on, but you cannot count on this to help you with your optimization efforts. There is evidence, however, that through the use of voice recognition technology, search engines are beginning to “hear” the content of videos, enabling them, in the future, to search through the audio content of a video to determine its subject matter and relevance to a particular search query.

There are three different scenarios for video optimization. It is worthwhile to do all three. Here are some factors relating to each scenario:

Video search engines that crawl your site

This is the scenario where you host the video on your own site. An example of a video search engine that will crawl your site to find videos is [Truveo](#). Although this type of search engine represents a small percentage of your total market opportunity, it is still worth pursuing. It provides a variety of optimization opportunities, including the following:

- This is normal HTML, so you can customize the content on the web page where the user can access or read about the video. Pay particular attention to the on-page text, as you can write a nice, keyword-rich, and compelling description of the video.
- You can keep your video files in one directory. This makes it easier for the crawler to find them all.
- If you embed a Flash video player into a web page, you can include a link to the raw video file (in AVI, WMV, QuickTime, etc.), to maximize exposure to all the different video search engines.
- You can use a Video Sitemap, which will make it easier for the crawlers to find your videos.

Video sites that accept RSS or MRSS feeds

In this scenario, you submit an RSS or MRSS feed to a third-party site. Examples of sites that accept such feeds are [Yahoo! Video](#) and [Blinkx](#). The basic approach to optimizing for this type of site is to enter a title, description, keywords, thumbnails, a category, and a full transcript. A full transcript is a good way to provide good fodder for a video search site to chew on. The transcript will provide more keywords that the search engine can find and choose to rank you for.

Video upload sites

These are video search engines that require or allow you to upload the video to them.

Examples include:

- [YouTube](#)
- [AOL Video](#)
- [MySpace](#)
- [Revver](#)
- [Metacafe](#)
- [Vimeo](#)
- [Flickr](#)
- [Dailymotion](#)
- [Yahoo! Video](#)

You can also use a site such as [TubeMogul](#), upload the data just once, and get some analytics data back as well. Also, make sure you keep your videos from getting too large. Videos smaller than 100 MB will work on most video sharing sites, and YouTube will

accommodate videos of up to 1 GB and up to 10 minutes in length (as of early 2009). Flickr limits the length of the videos it will support to 90 seconds.

There are opportunities to optimize these videos as well. Use the following tactics:

- For these sites, you will need to enter the metadata again when you get there. This will include a title, keywords, and a description. Also make sure to incorporate into your description a link back to related content and videos on your site, and YouTube will make the URL clickable provided it is a complete URL (including the http://). Make sure you are getting qualified traffic in return for providing video.
- Many publishers wonder what the best approach is for them. For example, the downside of posting your video on YouTube (or any other video sharing site) is that anyone can embed it on his site, and most of the links related to the video go back to the video sharing site page rather than your own site. These seem like strong disadvantages, and this is particularly true if you already have a site with lots of traffic on it.
- For many sites, though, the volume of traffic on YouTube in particular is compelling. As we noted in [Chapter 1](#), in January 2009 YouTube became the second largest search engine on the Web (passing Yahoo!). This is further enhanced by the advent of Universal Search in which the videos on YouTube (and found elsewhere by Google in its crawl of the Web) can get embedded in Google web search.
- For these sites, you should view a video strategy as an indirect link-building strategy. Use it to build some buzz about your site. YouTube is also a powerful social network. Getting YouTube viewers engaged with your videos can result in them spreading your videos around to other YouTube users in their network. YouTube offers many social features that will help a good video to spread virally, such as comments, video replies, favorites, ratings, and more.
- Make sure you include links back to your site so that people know where the video came from. Better still, have some content that complements and enhances what the user experienced in the video so that you are more likely to drive some links to your site as a result.

Other video optimization tips

There are other things to consider in optimizing videos:

- Create compelling content. If the content is not compelling, users will not benefit from it, and neither will your business. Note that even an instruction manual on how to use your left-handed bottle washer will be compelling to someone who needs the information, if you put it together correctly.
- Develop videos that reinforce the message of your brand. It is great to have a video go viral or be a success, but will it actually help you if it does?

- Make sure you include the word *video* on a regular basis throughout any metadata, as many user searches include the word and it will increase your chances of showing up in those queries. Of course, don't go hog-wild with it. Note that YouTube has a title tag display limit of 62 characters, so don't go beyond that.
- Pick a preferred format, such as *.mov*, *.avi*, or *.wmv*. There is no real win in offering the file in multiple formats from a video SEO perspective, but it will be useful to users who want to pick a preferred format. So, make just one of the versions visible to the search engines, and offer them all to users.
- Keep the length reasonable (five minutes or less). If the video is too long, you will run the risk of boring the user. There are exceptions to this rule, of course, but if you exceed five minutes be very certain that most viewers will be able to remain engaged.
- Give your video a keyword-rich but catchy title. The title is the biggest factor in convincing a user to view the video.
- Make good use of thumbnails to help users decide which video to watch. Users like to watch a quick snapshot so that they can decide whether they want to invest their time in the full video. Unfortunately, the way you generate thumbnails differs from site to site. But it is an important enough factor in getting people to view your video that it is worth going through the process for each site you upload to.
- Allow users to rate your video. Search engines will pay attention to this when ranking these videos. This is also a crucial part of going viral, as potential viewers are very attracted to highly ranked videos.
- Commercial content should include a watermark to indicate origin. This will help if your video goes viral and is copied to multiple sites, because it will still carry your branding message and/or site information.
- Use keywords in the filename and URL. Video search engines place a fair amount of weight on these in the world of video. Avoid using stop words, such as *the* and *and*, in the filename.
- Use appropriate keywords in the link text for the file. Keyword-rich anchor text carries a lot of weight.
- In preparation for voice-recognition-based indexing of video and audio content, develop video content "scripts" that are keyword-rich and relevant to your video's topic. In this regard, digital media SEO will be a very new and dynamic area of optimizing content for search engines.

Publicizing your video

Here are some basic tips that will assist in the promotion of your video:

- Put your video everywhere. Exposure is king. Upload the video to all the video upload sites previously identified. Be sure that you also reference your multimedia files within appropriate RSS or MRSS feeds.

- Use social bookmarking tools that will help make the videos more visible. This works regardless of promotional medium. Social bookmarking sites can help your video go viral.
- Offer the option to embed your video in other people's sites. This really encourages people to spread your video around.
- End the video with a "send to a friend" option. Give users some incentive to help you spread the word.

You can also find more video SEO resources at <http://www.reelseo.com>.

Conclusion

Optimizing for the vertical engines owned by the major search engines, as well as other players with small but growing market share, represents an opportunity that is largely untapped. You can get a leg up on your competition by optimizing your content for the appropriate engines. This may require enhancing or otherwise modifying the content itself, augmenting the content with additional metadata, or submitting the content by filling out forms or submitting feeds specific to each engine. All of this may seem like a lot of effort, but if you have the right type of content, it is worth it. Not only can you gain visibility in those vertical search properties, but you may also find that your listings are promoted into the blended search results in the main engines.

CHAPTER NINE

Tracking Results and Measuring Success

SINCE THE NATURE OF WHAT SEO PRACTITIONERS DO IS NOT CLEAR TO MANY PEOPLE, the expert practitioner knows how important it is to establish quantifiable goals and demonstrate results. Expectation management is key, and effective communication of timetables for deliverables and improvements is crucial to the success of any SEO project.

Web analytics and other data-mining tools can provide a rich array of data that is invaluable to the SEO process. This includes tools for discovering new links, finding long tail search terms, determining the best links your competitors' sites have, and more. All in all, a strong understanding of these tools will make you a better SEO practitioner.

However, thousands of tools are available, and it can be difficult to figure out which ones you need to use, which ones are merely nice to have, and which ones are not worth the time. It is a good idea to start simple and focus on what you need to establish a baseline measurement of where you are before you begin your SEO efforts and how to measure the results.

In other words, you need to measure whether you have been successful, and then get a sense as to how successful. The first thing you need to do is define why the site exists. Here are some possibilities:

- Generate sales
- Obtain leads
- Generate ad dollars
- Get people to read specific articles
- Obtain donations

These are all valid objectives that relate directly to the business. Missing from this list are things such as measuring rankings, measuring the total number of links, and measuring the average cost per acquired link (organically obtained links still have a cost: the cost of the marketing campaign to get them). These other measurements may be interesting, but they are not the end goal.

Closely connected with this are direct drivers of results. These include parameters such as:

- Total site traffic
- Traffic to your most popular pages
- Keywords driving the traffic
- Keywords driving conversions

The mechanics of measuring results in terms of the main business goal should always be the first set of measurements you put in place. From there you can diverge and look at other metrics that might help you diagnose problems with the site or give you additional insight into how to set up a campaign.

Why Measuring Success Is Essential to the SEO Process

Although quantifying deliverables and measuring progress are important for external reporting purposes, it is just as important for SEO practitioners to measure the efficacy of their own efforts to make timely adjustments as necessary. As you will see in this chapter, numerous tools are available to aid in this process.

At the beginning of any SEO project it is wise to establish baseline data points for the website. This includes the following:

- Quantifying organic search traffic by search engine and keyword
- Quantifying a baseline of the major keywords that are driving traffic by search engine
- Quantifying a breakout of what sections are getting the current organic search traffic by search engine and keyword
- Quantifying data on conversions broken down by search engine and keyword
- Identifying poorly performing pages
- Tracking search engine crawler activity on the site
- Determining the number of indexed pages
- Identifying 404 error pages and external sites linking to these pages, if any

Remember: you cannot methodically improve what you cannot measure.

Defining and mapping the path toward concrete goals are crucial aspects of the SEO process—and over time, some goals may change. Therefore, it is also important to make sure the data you capture helps you understand your progress against these goals.

In the world of web analytics, this is referred to as picking Actionable Key Performance Indicators (or Actionable KPIs). The best data measurements are those that potentially result in an action being taken in response. Think of this as data-driven decision making.

The Tracking Cycle: Produce, Launch, Measure, Refine

In summary, the basic process usually looks something like this:

1. Define an SEO campaign and set goals.

What are you going to accomplish, and what is the strategy for accomplishing it? How will you measure progress?

2. Discuss your strategy.

The marketing and business development teams are your allies here—you want to ensure that your SEO objectives are based on the overall business and site objectives, both long- and short-term.

3. Establish a baseline.

Now that you are about to start and you have decided how you are going to measure progress, establish a baseline by recording the current stats prior to beginning work. Make sure you don't get a false baseline due to seasonal factors or some other unusual event.

4. Proceed with your project.

Implement the new pages, the site changes, the link-building campaign, or whatever you have planned. Put it in place and execute.

5. Collect data.

Collect the newest data for each metric you decided to focus on. Since this is SEO and SEO can take many months to show results, make sure you wait long enough for your efforts to have an impact. Of course, if you are a student of the process, you can take more frequent measurements so that you can see how things begin to progress over time. For many on-page changes, 60 to 90 days are enough, but for link-building campaigns it may take six months or more to see the full impact. Many factors could influence the length of time you should wait. Here are some of them:

- If your site is brand new, it may take longer for your changes to take effect.
- If the scope of the change is drastic (such as a complete redesign), the time to see results will probably be longer.
- Sites that get crawled at great depth and frequency will probably yield results faster.
- Sites seen as authoritative may also show faster results.

6. Compare the baseline data to the new data.

The new data has little meaning unless it is compared to your baseline. This is the time when you can really assess your progress.

7. Refine your campaign.

Now that you have compared your old data with your new data, you can make some decisions. Is the campaign a bust? If so, abandon it and move on to the next one. The old business axiom “Fail quickly” applies here. The faster you diagnose a failure and move on to the next thing, the better.

You may also find you are getting mediocre results. Examining the data more closely may give you some ideas as to how you can improve those results. And if you are achieving great results, look for ways to scale the effort and drive even more volume.

Using Analytics As a Business Case for SEO

You can use a properly structured plan as the business case for an SEO project. The way to do this is to express the target results of an SEO project in terms of financial impact. You could include a variety of metrics in a business case, such as:

- Revenue
- Lead generation
- Margin
- Branding value
- Reach
- Other action triggers (newsletter sign-ups, contact requests, demo requests, free-trial acceptance, viewing a specific piece of content, etc.)

Measuring such things requires that you tie organic search engine visits to the revenue and other conversions that result.

Measuring Search Traffic

Classic web analytics data is an incredible asset to SEO. Here are three examples of ways to utilize this data for SEO purposes:

- Look at your daily referral reports to detect newly received inbound links (the great majority of webmasters click on a link after implementing it, to make sure it works)
- Look at the search terms people use to come to your site to spot long tail search opportunities
- Measure the results of your campaigns by tracking the increase in conversions you are driving over time

Web analytics are a must-have for any web publisher.

Basic Overview

Your hosting company most likely provides a free web analytics solution, such as [AWStats](#), [Webalizer](#), or something similar. Although these tools provide valuable data, they are very limited in scope, and other tools out there provide significantly more data. Here are six of the best-known ones:

- [Google Analytics](#)
- [Omniture](#)
- [Core Metrics](#)
- [Unica Affinium NetInsight](#)
- [Webtrends](#)
- [Yahoo! Web Analytics](#)

Web analytics platforms track your site's traffic in two major ways. The older of the two methodologies is to use software that analyzes your web server logfiles after traffic has occurred. Setting up this type of solution generally involves installing the software on an internal server that is able to gain access to the logfiles.

The newer methodology involves implementing JavaScript tags on all of the pages of your website. Provided that you have a basic template for your pages, this generally is a fairly straightforward process. JavaScript then tracks the activity on your web pages as it happens and builds a data profile reflecting that activity.

Selecting the Right Analytics Package

LogFile tracking and JavaScript tracking are equally valid methods, and each has its own strengths and weaknesses. The biggest advantage of the logfile method is that you can track search engine crawler activity on your site. This is something you cannot do in JavaScript implementations, because search engine crawlers do not execute the JavaScript.

The second big advantage of a logfile-based solution is that you run the software in-house, so no third party has a copy of a logfile with your proprietary traffic data on it. This distinction can be a big security issue for some organizations.

Ultimately, most companies opt for JavaScript tracking because JavaScript offers a much greater level of flexibility than logfiles. You can tweak the JavaScript to do custom conversion tracking, or gather pages into logical groupings in a manner that cannot be done in logfile-based applications.

Some companies, such as Unica and Webtrends, will offer you both options, or they will even offer the option to implement a combined solution. This kind of approach can bring you the flexibility and power of JavaScript, but still get you your search engine robot crawling data as well.

Making this decision is only the first step in picking an analytics package. We listed six of the more popular vendors earlier, and there are many more vendors than that. Each of these packages has different strengths and weaknesses. Not only that, they all do their counting a little bit differently. The chart from [Stone Temple Consulting's Web Analytics Shootout report](#), shown in Figure 9-1, helps to illustrate the point.

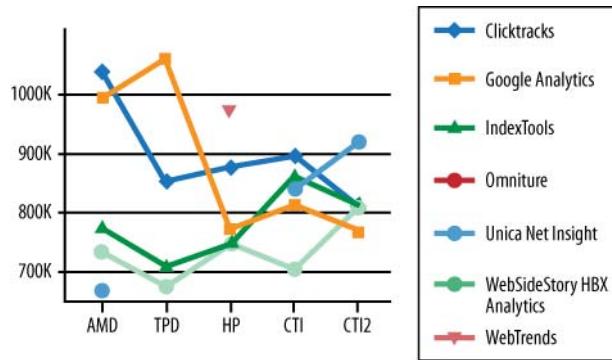


FIGURE 9-1. Web analytics accuracy

The chart in Figure 9-1 shows the results of seven different analytics packages (listed on the right). The traffic data for each vendor is reported across five different sites (represented by the acronyms on the bottom).

On the AMD site, the lowest reporting package (HBX Analytics) shows a little less than 750,000 unique visitors, and the highest reporting package (Clicktracks) shows about 1,050,000 unique visitors during the same period—almost 50% more!

These differences result from different decisions in how the analytics packages conduct visitor tracking. None of them are right or wrong, they are just different in the exact thing they are measuring.

NOTE

HBX Analytics, which is listed in Figure 9-1, is no longer available, as WebSideStory was acquired by Omniture. In addition, IndexTools was acquired by Yahoo! and is now called Yahoo! Web Analytics.

The more important component of this is whether the functionality of the web analytics software fits your needs. Making this more difficult to understand is the fact that you often do not know what your requirements are until you have used analytics for a while. As you engage with analytics, you will continually learn more things you want to investigate, and develop new requirements.

For many companies, one of the best solutions is to start with a free analytics package such as Google Analytics or Yahoo! Web Analytics and then look to buy a higher-end solution once they have pushed these packages to their limits. By pushing the limits of these free analytics products first, you will end up developing a set of requirements you can use in deciding where to go next.

None of this is meant to say that you should not brainstorm your requirements in detail before selecting an analytics package. You should (and must). Just expect that you will develop new requirements along the way. Web analytics is a journey that unfolds over time.

Based on the requirements you establish in your upfront brainstorming, you may find that you require a set of features that the free packages do not provide. Use that as knowledge to select the right package to start with.

Valuable SEO Data in Web Analytics

You can extract all kinds of data from web analytics. Here are a few of the more interesting types of information you may want to extract.

Traffic by search engine

One of the first things you may want to know is the breakout of traffic by search engine.

[Figure 9-2](#) provides an example of such a report in Google Analytics.

Notice how small the traffic is on Yahoo! compared to Google (about 2.5%). This may be indicative of a problem with the site in question and Yahoo!. This site owner might want to spend some time exploring why the site traffic from Yahoo! is so low.

Traffic by keyword

One of the basic data points of interest for an SEO practitioner is what search terms are bringing traffic to the website. This provides a quick way to see where the SEO campaign is going well and where it is not going so well. You can also use this to spot opportunities where a key search term is providing some traffic, but not as much as you would expect if you were ranking highly for that term.

You can then look to see where you are ranking for that term. Perhaps you are in a lower position on the first page, or on the second page of the SERPs. If so, it might make sense to focus some attention on this term. With a little effort, such as a focused link-building campaign for the page in question, you may be able to move up several positions and obtain a traffic boost.

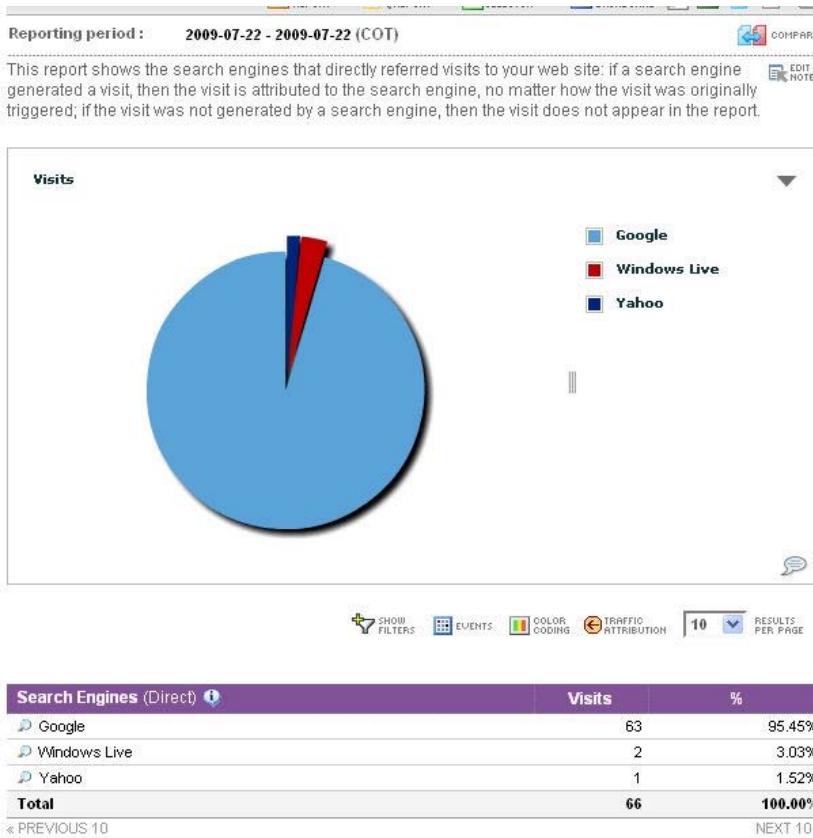


FIGURE 9-2. Traffic by search engine

A traffic-by-keyword report can also show you the long tail of search as it relates to your current site. Figure 9-3 depicts a snippet from the organic search phrases report of Yahoo! Web Analytics showing some of the small-volume terms for [Stone Temple Consulting's website](#).

Notice that a lot of names are showing up in the list, in addition to the rather interesting *who is the author of ask.com*. Looking into this query data can give you a broad perspective on opportunities for long tail search.

Notice also how two of the names combine *Google* with a person's name (*jack ancone google* and *google carter*). This could indicate an opportunity to make sure the person's organization name is a part of the title of the articles in the [Stone Temple Consulting interview series](#).

andrew tomkins	2	0.05%
cross channel integration	2	0.05%
rajat mukherjee google	2	0.05%
google interview questions wiki	2	0.05%
podcast with transcript	2	0.05%
css cascading style sheet improve google search result	2	0.05%
adam lasnik google	2	0.05%
canonical redirects	2	0.05%
john ways müller	2	0.05%
kirsten rasanen	2	0.05%
seo idea	2	0.05%
marketing web analytics	2	0.05%
podcast transcript	2	0.05%
click fraud metrics	2	0.05%
grad conn seattle	2	0.05%
google analytics cookies	2	0.05%
custom video maker	2	0.05%
cse google	2	0.05%
ms natan	2	0.05%
subscribed link	2	0.05%

FIGURE 9-3. Long tail keywords

Segmenting Search Traffic with Multiple Parameters

Next, you can consider putting these things together. Even the free tools provide substantial capability for building out custom reports. Figure 9-4 depicts a screen shot from Yahoo! Web Analytics that shows the search phrases just for Google, and the pages of the site that Google sent users to when users clicked on your link.

This is a key improvement to your SEO research for two reasons:

- If you are looking into what terms can bring fast traffic benefits as a result of some additional optimization, you are going to want to know in which search engine you are ranking.
- If you are going to optimize a page to rank higher, you will need to make sure you are optimizing the right page!

Referring Sites

It is interesting to look at a referring site report for a number of reasons, but one of the more interesting SEO reasons to do so is to spot when you receive new links. You can often see those new links in these reports first, even before the search engines report them. Figure 9-5 shows a sample portion of the referring sites report from Google Analytics.

stone temple seo	4	0.11%
Google	4	100.00%
mails ohne	4	0.11%
Google	4	100.00%
frazier miller	4	0.11%
Google	4	100.00%
white hat link building	4	0.11%
Google	4	100.00%
css menu seo	4	0.11%
Google	4	100.00%
meta tags for seo	4	0.11%
Google	4	100.00%
how can css be used for developing a unique website style?	4	0.11%
Google	4	100.00%
technical seo	4	0.11%
Google	4	100.00%
enquisite	4	0.11%
Google	4	100.00%

FIGURE 9-4. Search phrases from one search engine

121.	 sitweb.ro	3
122.	 stackoverflow.com	3
123.	 stephanspencer.com	3
124.	 submitsuite.ro	3
125.	 tech.groups.yahoo.com	3
126.	 translate.google.cn	3
127.	 us.mg1.mail.yahoo.com	3
128.	 us.mg4.mail.yahoo.com	3
129.	 winningtheweb.com	3
130.	 1986x.net	2
131.	 360digest.com	2
132.	 backtype.com	2
133.	 briansolis.com	2
134.	 buddha.voce.ro	2

FIGURE 9-5. Referring sites report

Expanding on this example, consider the site circled in Figure 9-5, edwardbeckett.com. If this is the first time you have ever noticed this site in your referrers, it can be a leading indicator that

you have received a new link. This is of interest as it can help you to detect new links that result from your link-building campaigns, and therefore help you measure which of your link-building campaigns are yielding the best results.

Using Analytics Dashboards

In analytics terms, a *dashboard* is a single-page view that contains your most critical metrics all in one place. Of course, your most critical metrics are different from those of the next publisher, because the needs of different sites vary greatly.

In addition, multiple dashboards may be required in any given organization. For example, the CEO of a large public company probably wants to see different data (and a lot less of it) than a senior business analyst.

Each analytics package provides methods for implementing a custom dashboard. [Figure 9-6](#) is an example of one from Unica's Affinium NetInsight.

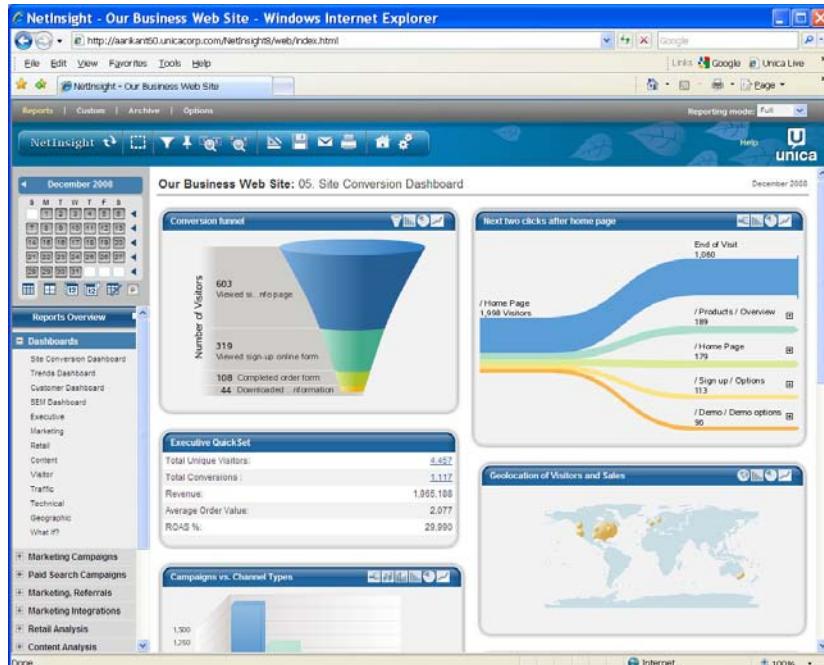


FIGURE 9-6. Custom dashboard

As you can see from [Figure 9-6](#), a dashboard can be quite visual. What is most important, though, is that it provides the data that is most important to the person for whom the dashboard

was designed. As an SEO practitioner, you can implement a dashboard to show progress against the goals you set for your SEO campaign.

Providing this type of visibility has two important benefits:

- The person viewing the report will appreciate the fact that she does not have to work hard to do a quick health check on the progress of the SEO efforts. As we suggested at the beginning of this chapter, the fact that you agree to measurable goals will be a great comfort to management.
- You will know what data your managers are looking at. When something goes wrong (or right), or when management wants to discuss some aspect of the business, they will have started from the dashboard you set up for them.

A Deeper Look at Action Tracking

Action tracking is one step deeper than basic analytics. Rather than simply observing what pages are visited and how many unique sessions are logged, action tracking allows you to narrow down groups of visitors based on the actions they take on your site.

In most instances, it requires setting up a code in your analytics program and attaching that code to a button, page load, image rollover, or other JavaScript-trackable task (a click or hover). Once you've plugged it into your analytics and the website, you can use the action to refine data you're already collecting. [Figure 9-7](#) provides a look at how this works.

You can see from [Figure 9-7](#) that:

- SEOMoz's sign-up form has action tracking applied to it.
- Based on the people who sign up, you can predict which search terms will be better at converting visitors into applicants.
- The Revenue column is empty, but if you were tracking e-commerce buyers, you could put their totals into the Revenue column and track high-volume buyers.
- Expanding on this idea, you could also track users by time of day, the search engine they used, their geographic location, and so on.

So, what types of actions should you be tracking on your site? The answer varies depending on your business and site structure. Here are some suggestions as segmented by site type.

E-commerce site:

Add to Cart button

Studies have shown us that users who "add to cart," even if they do not complete the checkout process, are more likely to return to make a purchase. This is also a good way to calculate shopping cart abandonment and make changes to refine and improve the process.



The figure shows a screenshot of a user sign-up form titled 'USER SIGNUP' from eGifter.com. At the top, there are fields for 'Email' (e.g., 'egeorge@stardust.com') and 'Password' (with a 'Forgot your password?' link). Below these are fields for 'First Name' and 'Last Name' (both with 'Required' validation messages), 'Email Address' (with 'Your email is already registered.' message), 'Credit Card Number' (with 'Card number required.' message), and 'Credit Card Expiry' (with 'Expiry date required.' message). To the right of these fields is a 'Why should I register?' section with a question mark icon. It explains that by registering, users get access to personalized pricing, can view their gift history, and receive special offers. There are also sections for 'Why do we ask?' and 'What is the MyGifts 101 Knowledge Center?' with similar explanatory text. At the bottom of the form is a 'REGISTER MY ACCOUNT' button.

FIGURE 9-7. Action tracking in analytics

Complete checkout

An obvious one; this action will show you what percentage of each user group is converting into sales.

Save to wish list

E-commerce sites offering wish lists are still in the minority, but wish lists are a great way to track interest that isn't quite a purchase.

Send this to a friend

Many sites offer a “share this page” function, and it is a great action to be aware of. If folks are sending out your link, you know you have a hit.

B2B site:

Subscribe to newsletter

A subscription is a tacit endorsement of your brand and a desire to stay in contact. It may not be a conversion, but for B2B, it may be the next best thing.

Contact form submission

A runner-up with subscribing to a newsletter is filling out the contact form. Though some of these forms will report support issues, many may contain questions about your products/services and will indicate a desire to open a sales conversation.

Email link

As with contact forms, direct email links have the possibility of becoming a sales contact. The best thing you can do is clearly label sales emails and track them separately from support or business issues.

Blog:

Subscribe to RSS feed

An RSS feed subscriber is a blog's equivalent of a conversion; tracking these is imperative.

Add comment

Anyone who is contributing content to the blog or participating should be paid attention to (as should those channels that send you participatory people).

Social bookmark/share

All those folks who are submitting your content to Digg and Reddit deserve to be recognized (and sought after).

Forum or other site based on user-generated content (UGC):

Sign up for an account

These users are active contributors; you need to know where they come from.

Contribute content

When a user publishes, discovering her path is important (especially if it is not from a bookmark/type-in).

Add comment

As in the preceding list item, comments are a great predictor of engagement.

Vote/rate

Even low levels of participation, such as a rating or a vote, are worth tracking when every piece of participation counts.

You can get very creative with many of these and track all sorts of actions. If you offer a tool or calculator, track its usage. If you want to follow who clicks a particular link (rather than just a page), add an action to it. You can even see which users hover on an image.

For many marketers, action tracking is the secret sauce. Once you install and activate this data, it is easy to adjust SEO (or PPC) marketing efforts on particular terms, phrases, or pages that will make a huge difference to the campaign's efficacy.

For SEO professionals, action tracking is a “cannot-live-without-it” tool. The recommendations made based on this kind of tracking are backed up by real data. High-volume action-tracking studies almost always make a huge impact on a site’s bottom line.

Separating the Analytics Wheat from the Chaff

One of the big problems with web analytics is that there is so much data. There is seemingly an endless supply of the stuff—so much that you can get lost in the weeds and waste a lot of time. As Yahoo! Web Analytics’ Dennis Mortensen (<http://visualrevenue.com/blog/>) likes to say, do not “report-surf.”

By this he means you shouldn’t get lost in the data, looking at one interesting fact after another, without understanding what business objective you are attempting to accomplish. To avoid this trap, you need to have a plan.

The basis of this plan is the definition of actionable Key Performance Indicators (KPIs). To summarize this concept very briefly, do not view a report unless there is a chance that it will provide enough insight to take action (such as reallocating resources to fix a problem or pursue a new opportunity you discovered).

Common analytics mistakes

Because of the complexity of web analytics, there are plenty of opportunities to make mistakes. Even if you are laser-focused on actionable KPIs, you can still draw wrong conclusions from the data. Here are a couple of examples:

Making a decision on too small a sample size

A simplistic example would be if you had a keyword that brought you two visitors, and both of them bought something. What does that mean? Your conversion rate will always be 100%? Of course not.

You need to have a test run for a length of time that is statistically significant. According to [Jonathan Mendez](#), long-time evangelist for Offermatica (which is now part of Omniture’s Test & Target product), the “golden rule” for test length is 100 conversions per branch (or tested element). For an A/B test, ideally you would have 200 conversions before you looked at the results. However, if after 50 conversions you have 40 on one branch and 10 on the other, your test is over. If your variance in results is lower, you need more data to make sure the data you are seeing is accurate.

Not recognizing when you have enough data

You can make this mistake in the opposite direction too. When you have enough data, you need to be willing to act on it.

Working with biased data

You can easily collect biased data and make the wrong decision. For example, an e-commerce site basing its SEO plans on December's traffic numbers is probably being misled. And temporal factors aren't the only cause—visitor locations, referral sources, and in-house campaigns to existing lists can all skew information.

Tying SEO to Conversion and ROI

As we discussed previously in this chapter, it is important to tie your SEO campaign to the results it brings to the business. A fundamental piece of that is measuring the conversions driven by organic SEO traffic. Here are some of the most common types of conversions:

Sales/sales revenue

This is the one that everyone assumes is part of conversions. Sales and sales revenue (or better still, margin) conversions can be the simplest things to track, except when you are selling many different products at different price points and in different quantities. In this case, the process would need to be a bit more sophisticated.

If your site is advertising-driven, you need to look at the impact of organic search traffic on advertising revenue. If you have no financial goals for your site, you need to look at some of the other types of conversions.

Email/blog/newsletter subscriptions

Anytime a user signs up to receive regular communications from you it is a win. Even though there are not direct financial consequences to this it is still a conversion. Someone who has subscribed to something you offer is more likely to become a customer than a first-time visitor to your site, so you need to credit this type of conversion.

Sign-ups

Closely related to the notion of subscriptions are other types of sign-ups. Perhaps you offer a service such as a tool that people need to sign up for to use. Even if the tool is free, you should track this as a conversion.

You most likely received the person's email address in the process, and even if she indicates that she does not want to receive commercial communications from you, you should be building loyalty with the tool you provided her access to (or whatever she signed up for); otherwise, why would you be providing her that service?

Downloads

Many sites offer free downloads, such as white papers, or free downloadable tools. Even if you do not require a sign-up of any type, you should still count this as a conversion.

You are getting your message out there with the downloads you offer.

Contact forms and phone calls

This is when someone requests that you contact her, or she contacts you.

Visitors who share

This conversion happens when a visitor shares the information she found on your site with someone else. For example, if your site has a “share with a friend” or an “email to friend” feature, you could keep track of conversions by noting each time a user used that feature.

Visitors who link

A user who links will visit your site and find its content useful, entertaining, or otherwise compelling enough to link to it from her own site.

Visitors who publicize

Visitors can publicize your site by mentioning it in forums, blog comments, or social media sites such as Twitter and Facebook, or by writing about it on their own site.

You should place a dollar value on every type of conversion you receive. Even a white paper download has some value to it. Perhaps it may increase the potential of a future sale to that user. For example, if 100 people download your white paper and two of them become customers and buy a \$100,000 consulting contract, you can estimate the value per download at \$2,000.

Attribution

Another issue to be aware of is that of attribution. The issue is that there is a tremendous amount of interaction between types of media. Here are some examples:

- A user does a search, clicks on an organic search result, and then reads a few things and leaves. The next day, she remembers what she read, does another search, and this time clicks on a paid search ad for the site and buys a product (organic search should receive some credit).
- A user does a search, clicks on an organic search result, and leaves. A few days later, because of what she learned, she goes into a store and buys a product (organic search should receive some credit).
- It goes in the other direction too: a user sees a TV ad, then does an organic search to find the website, and buys the product (the TV ad should get some credit for the sale!).

Such cross-channel interactions are becoming increasingly common. Recent data published on [Search Engine Watch](#) supports this. For example, [HealthPricer.com](#) reported that about 6.5% of its transactions took place at least one day later than the customer’s original visit to its site.

Similarly, a ScanAlert study of a large number of e-commerce sites showed that 30% of transactions took place one to three days after the customer’s initial visit. So, the problem of attribution management is a significant issue!

It is not always easy to track multivisit conversions in detail. Certainly, a direct interaction between TV/offline stores and a website is very difficult to track in detail. But even looking at

multiple site visits with an interaction between PPC and organic search can be difficult, partly because the tracking systems rely on cookies, which users have a tendency to delete.

So, attribution is something you should be concerned about, and ideally you should attempt to track the impact across multiple site visits. However, the methods for attributing sales across multiple site visits are imperfect. One thing you can do to minimize the impact of that is to count different types of conversions, such as filling out Contact Us forms, downloading a white paper, visiting a key product page, and so on. Tracking a larger number of events that are indicators of progress toward a sale (or whatever your end goal may be) can help you identify important helper keywords that are playing a key role in generating results further down the line.

Setting Up Analytics Software to Track Conversions

One of the great things about tracking conversions is that you can use the resulting data to focus your efforts. If a given keyword, page, or referrer is showing much higher conversion rates than another source, you may want to focus more of your SEO efforts on those rather than on something else that does not convert as well.

Capturing conversion data takes a bit of setup. Your analytics software does not know what you consider a conversion unless you define it. For example, in Google Analytics, you need to go into your Profile Settings to configure conversion goals and funnels (see [Figure 9-8](#)).

Conversion Goals and Funnel <small>?</small>				
Select up to 4 conversion goals for this profile, and define the funnel pages leading up to each goal.				
Goal Name	URL	Active Goal <small>?</small>	Settings	
G1 Contact Requests	http://www.stonetemple.com/STC_Contact.shtml	On	Edit	
G2 (Goal not configured)		Off	Edit	
G3 (Goal not configured)		Off	Edit	
G4 (Goal not configured)		Off	Edit	

FIGURE 9-8. Google Analytics goals

Once you click Edit (on the right) for one of the goals, you come to a screen that allows you to define a goal (see [Figure 9-9](#)).

In [Figure 9-9](#), the primary data being tracked is when someone arrives at a given web page. These types of conversions are easy to set up.

Event tracking, such as someone clicking on a link, is slightly more complicated. Since you cannot count on a page load to tell you when a link has been clicked on, you need to add a bit of tracking code to the link itself. Here is what the tracking code looks like for Yahoo! Web Analytics:

```
<a href="http://www.yourdomain.com/bigmoneypage.html" onClick="ACTION='05'">
Spend Big Money Now</a>
```

Goal Settings: G2

Enter Goal Information

Active Goal: On Off

Match Type : Regular Expression Match

Goal URL : (e.g. For the goal page "http://www.mysite.com/thankyou.html" enter "thankyou.html")
To help you verify that your goal URL is set up correctly, please see the tips here.

Goal name: Goal name will appear in Conversion reports.

Case sensitive URLs entered above must exactly match the capitalization of visited URLs.

Goal value: 0.0

Define Funnel (optional)

A funnel is a series of pages leading up to the Goal URL. For example, the funnel may include steps in your checkout process that lead to the thank you page (goal).

Please note that the funnels and 'Required step' that you've defined here only apply to the Funnel Visualization Report.

Note: URL should not contain the domain (e.g. For a step page "http://www.mysite.com/step1.html" enter "/step1.html").

	URL (e.g. "/step1.htm")	Name	<input type="checkbox"/> Required step
Step 1	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>
Step 2	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>
Step 3	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>
Step 4	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>
Step 5	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>
Step 6	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>
Step 7	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>
Step 8	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>
Step 9	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>
Step 10	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>
Goal (use above)	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>

Save Changes **Cancel**

FIGURE 9-9. Google Analytics goals setup

The key code is `onClick="ACTION='05'`. This tells Yahoo! Web Analytics to associate this event with Action #5. This requires some additional configuration within Yahoo! Web Analytics so that you can give Action #5 a label and a value, as shown in [Figure 9-10](#).

As you can see in [Figure 9-10](#), actions 1 through 4 are already set up. Now all you need to do is give Action #5 a label.

Form submissions are another thing you may want to track. This could include when someone has completed a request to sign up for a newsletter, requested or started a white paper download, or simply used a form to request that you contact her.

These are usually quite easy, provided that the result of the form submission is some sort of “thank you” page. All you need to do in this case is tell your analytics software that the loading of that “thank you” page is considered a conversion, and then give it a value.

Life gets more complicated when you want to track actual product sales on sites where products vary in price. Generally speaking, this will require putting some custom JavaScript on your order confirmation page.

YAHOO! WEB ANALYTICS

The screenshot shows the Yahoo! Web Analytics interface. At the top, there's a purple header bar with links for REPORTS, CONTROL CENTER, INSTALLATION, SETTINGS, and LOGOUT. Below the header, the main content area has a title 'Custom Actions'. On the left, there's a sidebar with a link to 'Back to settings...'. The main content area contains a sub-section titled '1. Name Your Actions' with a note about customizing descriptions. It also includes a note about actions being tracked multiple times by default. A table lists 13 actions, each with a checkbox for 'Unique' tracking.

Action ID	Action	Unique
Action #1	SALE	<input type="checkbox"/>
Action #2	SIGNUP	<input type="checkbox"/>
Action #3	SUBSCRIPTION	<input type="checkbox"/>
Action #4	CONTACT FORM	<input type="checkbox"/>
Action #5	Action #5	<input type="checkbox"/>
Action #6	Action #6	<input type="checkbox"/>
Action #7	Action #7	<input type="checkbox"/>
Action #8	Action #8	<input type="checkbox"/>
Action #9	Action #9	<input type="checkbox"/>
Action #10	Action #10	<input type="checkbox"/>
Action #11	Action #11	<input type="checkbox"/>
Action #12	Action #12	<input type="checkbox"/>
Action #13	Action #13	<input type="checkbox"/>

FIGURE 9-10. Yahoo! Web Analytics action setup

In addition, you will need to have your web application populate certain data in the JavaScript before executing it, such as the total amount spent. Oftentimes, publishers choose to include a lot more data, such as listing all the products sold, the price per product, how many units, shipping costs, and so on.

Although this is a lot more complex than the other scenarios we've outlined, it is still achievable. Here is a link to the Google page that describes how to set this up for Google Analytics: <http://www.google.com/support/analytics/bin/answer.py?hl=en&answer=55528>.

Although the preceding examples used Google Analytics and Yahoo! Web Analytics, any of the major analytics packages can do this type of conversion tracking.

Conversion tracking strategy

An important step in conversion tracking is deciding what you want to call a conversion. As we outlined earlier in this chapter, it is important to consider all types of conversions, such as

sign-ups, contact requests, downloads, and so forth. There is one other decision you need to make: how to value each conversion.

What is a contact request worth? What about a newsletter sign-up? One way to back into this is to look at how many of your newsletter sign-ups become customers. For example, if 100 people sign up for the newsletter, 10 of them become customers, and the total dollar value of their orders is \$200, the average value per sign-up is \$2.

When you first set up conversion tracking you may need to estimate some of these things. But as your data improves, you can improve your calculation and refine the value you place on each conversion.

Segmenting Campaigns and SEO Efforts by Conversion Rate

Now that you have conversion tracking set up, how do you use it to focus your SEO efforts? One of the things you will want to do is to track conversion data in different ways:

Conversions by keyword

What keywords are bringing the best results?

Conversions by referrer

Which traffic source is offering the best conversion?

Conversions by web page

Which pages on your site result in the highest conversion?

Conversions by initial entry page

Which initial entry pages drive the most conversions?

Conversions by keyword and landing page

Taking into account the highest-converting landing pages on your site, what keywords are the big hitters for those pages?

For example, SEOMoz actively tracks its conversions using Yahoo! Web Analytics. Tracking by keyword can be a useful way to do this, as shown in [Figure 9-11](#).

You can see that the conversion rate on keyword-tool-related keywords is high. Perhaps an SEO campaign to improve the ranking of SEOMoz on these keywords is in order. Not only do you see high conversion, but you also see reasonable material volume.

Implementing an SEO campaign that doubles related traffic for this set of keywords will likely have a significant impact on the bottom line.

Increasing Conversion

As an SEO practitioner, you should strive to become an expert at conversion optimization, because higher conversion rates mean higher impact for your SEO campaigns. (Co-author

Search Phrases (Direct)	First Time	Returning	Visits	Action	Revenue	Conversion
seomoz	2,687	2,438	5,125	134	\$0.00	2.61%
seo tools	704	178	882	58	\$0.00	6.58%
free seo tools	172	39	211	28	\$0.00	13.27%
seo moz	611	607	1,218	25	\$0.00	2.05%
seo	1,291	312	1,603	21	\$0.00	1.31%
seomoz.org	301	134	435	21	\$0.00	4.83%
seo test	188	35	223	18	\$0.00	8.07%
seo blog	513	191	704	9	\$0.00	1.28%
seo tool	80	15	95	7	\$0.00	7.37%
trifecta	35	17	52	7	\$0.00	13.46%
Subtotal	6,582	3,966	10,548	328	\$0.00	3.11%
Total	79,660	27,283	106,943	831	\$0.00	0.78%

[« PREVIOUS 10](#)[NEXT 10 »](#)

Tip: To add an alert, right-click the metric figure you wish to monitor.

FIGURE 9-11. Yahoo! Web Analytics tracking conversions

Stephan Spencer offers a viewpoint in MultiChannel Merchant at <http://multichannelmerchant.com/webchannel/0701-understand-users-needs/>.)

Nonetheless, the average site conversion rate for a typical merchant today is only about 2.4%. That's pretty depressing when you think about it.

So, how do you help your conversion rate by managing your content and refining your site design? Most web resources that give content advice are focused on blogs; specifically, what and how often you should publish. Those of us who use websites strictly for sales and marketing are equally concerned with good content, but from a totally different angle—one that traditional content advice doesn't properly address.

What's more, much of the design advice out there seems contradictory. The key to enhancing your conversions by adjusting your design and content is not in thinking strictly about the technology or the mechanics of site layout, but in understanding the needs of your visitors. In that, there are many different elements to consider. A book about SEO can't cover this topic in depth, but you may want to pick up one of the following excellent books on the topic: *Always Be Testing* by Bryan Eisenberg and John Quarto-vonTivadar (Sybex), *Landing Page Optimization* by Tim Ash (Sybex), or *Web Design for ROI* by Lance Loveday and Sandra Niehaus (New Riders Press).

The link bait bump

A specific example of how to use action tracking by referral source is to track a *link bait bump*. The idea is that after successful link bait draws thousands of visitors to your site, a small percentage of those visitors will stick around and continue to visit on a semiregular basis (this is particularly true if you're delivering fresh, valuable content on a regular basis).

Action tracking by referral source. Figure 9-12 shows a sample report from Yahoo! Web Analytics (which has been pulled after an article for SEOMoz made the home page of Digg).

Referring Domain (Direct)	First Time	Returning	Visits	Action	Revenue	Conversion
Direct access or bookmark	85,879	110,306	196,185	1,189	\$0.00	0.61%
http://www.google.com	44,491	17,621	62,112	454	\$0.00	0.73%
http://www.seomoz.org	4,500	22,671	27,171	505	\$0.00	1.86%
http://www.google.co.uk	6,509	3,895	10,404	90	\$0.00	0.87%
http://twitter.com	4,380	2,112	6,492	22	\$0.00	0.34%
http://www.google.co.in	5,417	613	6,030	44	\$0.00	0.73%
http://news.ycombinator.com	4,336	791	5,127	2	\$0.00	0.04%
http://www.stumbleupon.com	4,041	695	4,736	2	\$0.00	0.04%
http://www.google.ca	2,788	843	3,631	17	\$0.00	0.47%
http://search.yahoo.com	2,178	891	3,069	28	\$0.00	0.91%
Subtotal	164,519	160,438	324,957	2,353	\$0.00	0.72%
Total	244,136	180,813	424,949	3,389	\$0.00	0.80%

[« PREVIOUS 10](#)[NEXT 10 »](#)

Tip: To add an alert, right-click the metric figure you wish to monitor.

FIGURE 9-12. Action tracking by referral source

In [Figure 9-12](#), you can see a list of domains that sent SEOmoz traffic over a few weeks' time, along with the number of resulting SEOmoz membership sign-ups. You can see some clear patterns, including the following:

- Digg, StumbleUpon, and Reddit are sending traffic, but less than one-tenth of 1% of those visitors sign up for membership on the site.
- The domains that point primarily to the SEOmoz tools (such as SEOCompany and WebRankInfo) are responsible for the highest number of sign-ups, but this is to be expected since one has to register to use Keyword Difficulty in particular.
- SearchEngineWatch, Stuntdubl, and SEOBook are amazingly on-topic sources of traffic, which directly bears out in the higher percentages of referrals who become members.

Action tracking is one of the most valuable things you can do on any type of site. Most folks limit its application to e-commerce, but you can see from the data in [Figure 9-12](#) how valuable it can be, even for a simple blog. Even if you do not have a membership sign-up, you could attach the action tracking to your “add comment” link in a blog or to a newsletter sign-up on a content site. Segmenting that data by incoming referrals (search engines, direct type-ins, and links) gives you an even better picture of who’s sending you truly valuable traffic. You can also get an indication of traffic quality by looking at bounce rate, time on site, and number of pages viewed for each referrer.

In [Figure 9-13](#), you can see that around August 2006, SEOmoz published some successful link bait that helped bump up the site visitor count until October, when there was another big spike. This traffic “bump” after the spike is another great benefit of creating viral-worthy content, and is a terrific way to grow the popularity of a site.

The tricky part is to continue to deliver great content after a viral piece and maintain the creative energy necessary to prove to visitors that you’re worth coming back to again and again.

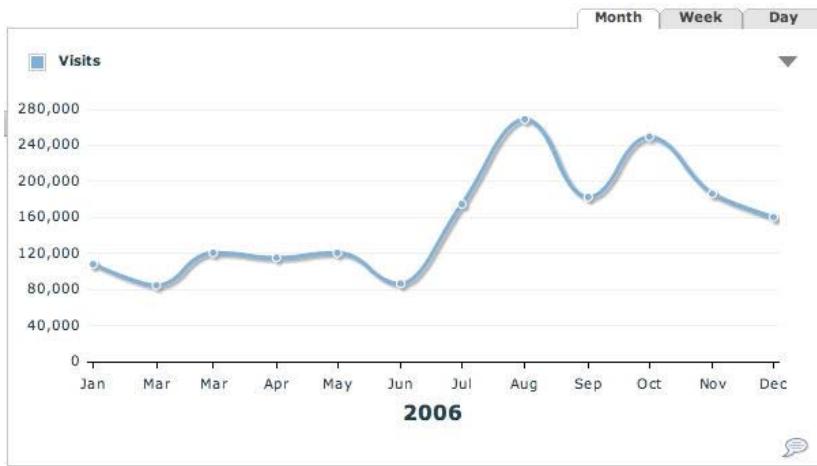


FIGURE 9-13. Traffic report

It is not an easy sell—people’s time is valuable. You have to make your content so targeted, interesting, and worthy of their attention that they cannot look away. It is challenging and time-consuming, but it is a great way to build a successful site.

Determining Project ROI

An effective SEO process is one that continuously works toward a positive return on investment. A basic SEO ROI business case would include these elements:

Number of people searching for your keywords

This can be hard to estimate, because you cannot completely map out the long tail. One rough estimation strategy is to multiply the search volume for the top terms for your business by 3.3 (i.e., assume that the head terms are about 30% of the available volume).

Expected click-through rate from organic search

This relates to the position you think you can achieve (refer back to [Figure 1-11](#) in [Chapter 1](#)). The #1 position can net you approximately 40% of the available clicks.

Average conversion rate

Once you have captured the visitor, how successful are you at completing a conversion? Typical conversion rates for a website might be between 2% and 5%. It should be easy to get this data from your analytics. You should already know what your conversion rate is!

Average transaction amount

Last but not least, factor in the average transaction value. Again, this is data you already have.

Ultimately, the formulas look like this:

SEO Revenue = People searching for your keywords * click-through rate * average conversion rate * average transaction amount

For example: 10,000 per day * 10% * 5% * \$100 = \$3,000 per day

SEO ROI = SEO Revenue / SEO Cost (use total \$ spent for salaries and consulting, or number of hours spent)

For example: \$3,000 per day / \$500 per day = an ROI of 6X.

Competitive and Diagnostic Search Metrics

Thus far in this chapter we have focused on the basics—the dollars and cents of determining whether you have succeeded in your SEO campaign. As we noted at the beginning of the chapter, these should be your first priority in your analytics efforts. In this section, we will start looking at metrics that you can use to diagnose specific SEO issues. An example of this would be finding out whether a major section of your site is not indexed. Another example is seeing how your traffic growth compares to that of your competitors (this helps you decide whether you have set the right objectives for your efforts).

Search Engine and Competitive Metrics

Numerous tools allow you to monitor your site and those of your competitors, providing insight into your SEO progress. You can also use these tools to figure out what your competitors are doing from an SEO perspective. This type of intelligence can provide you with new ideas on how to adapt your strategy to get better results.

As with all such tools, it is always important to understand the context of the tools and to have an idea as to what you are looking for. Better knowledge of your competitors' strategy is certainly one valuable goal. Detecting a problem in how your website is crawled is another.

By selecting specific and actionable goals, you can set yourself up for the highest possible return.

Site Indexing Data

It is valuable to know how many pages in your site are in a search engine's index. This is interesting for two reasons:

- To determine whether important parts of your site are not in the index. If key parts of the site are not in the index, you can embark on an effort to determine why.
- To learn about your competitors' sites and strategies.

You can get basic information on the number of indexed pages for a site using the `site:yourdomain.com` command in Google, Yahoo!, or Bing. [Figure 9-14](#) uses the results for Dennis Mortensen's blog.

Google search results for `site:visualrevenue.com`:

- VisualRevenue | Web Analytics & Online Marketing**
I believe you might be looking for the Web Analytics and Online Marketing blog of Dennis R. Mortensen...
visualrevenue.com/ - Cached - Similar - ⌂ ⌂ ⌂
- VisualRevenue | Web Analytics & Online Marketing blog**
We create a setup where we use Web Analytics Metric Alerts to illuminate a set of organic Search Phrase Opportunities. There is a whole industry dedicated ...
visualrevenue.com/blog/ - Cached - Similar - ⌂ ⌂ ⌂
- VisualRevenue | Privacy Policy**
VisualRevenue does NOT rent, sell, or share personal information about you with other people or non-affiliated companies except to provide products or ...
visualrevenue.com/blog/privacy - Cached - Similar - ⌂ ⌂ ⌂
- VisualRevenue | About Dennis R. Mortensen**
Hi there, welcome to ME! :) I regularly speak at conferences and corporate events, and conducts webinars and workshops. Feel free to connect or contact me ...
visualrevenue.com/blog/about - Cached - Similar - ⌂ ⌂ ⌂
- VisualRevenue | 2007 March**
Why? There is a revenue opportunity using competitive intelligence! I think it's naive to believe that you control your own web setting. ...
visualrevenue.com/blog/2007/03 - Cached - Similar - ⌂ ⌂ ⌂

FIGURE 9-14. Site indexing data

As shown in [Figure 9-14](#), Google reports 566 pages in the index. Is that all of his pages? Many site owners will have a pretty good idea of the number of pages on their site and therefore can get a gut feeling as to whether they have a problem. One important thing to note is that the results that `site:` commands return are not that accurate. You can use them to give you a gut-feel sense, but that is all.

You can improve this data by performing your site query and then appending `"&start=990&filter=0"` to the end of the Google SERP URL. This will give you more accurate totals for the number of pages indexed.

If you think many pages are missing from the index, you can begin to explore a bit deeper. You can start by narrowing down sections of the site. [Figure 9-15](#) shows the 2007 blog post archives on Dennis's site.

As shown in [Figure 9-15](#), this reports 87 pages in the Google index. By using these different command structures, you can isolate what portions of the site appear to be less well indexed than others. You can also combine the `site:` command with other operators, as we outlined in [Chapter 2](#). For example, consider combining it with the `inurl` operator, as in [Figure 9-16](#).

[Figure 9-16](#) shows that there are 8,840 pages of the `xbox360` subdomain on the <http://www.iup.com> website in the index.

So far, we've shown website indexing tools that you can use on any website, which is great for competitive research. However, you can pull more accurate statistics for your website using Bing [Webmaster Tools](#). You can also pull site indexing data from [Google Webmaster Tools](#) (Google WMT), but this tool is limited to reporting against pages that appear in your Sitemaps file.

The screenshot shows a Google search results page with the following details:

- Search Query:** site:visualrevenue.com/blog/2007
- Number of Results:** 1 - 10 of about 87
- Results:**
 - VisualRevenue | 2007 March**: Why? There is a revenue opportunity using competitive intelligence! I think it's naive to believe that you control your own web setting ... visualrevenue.com/blog/2007/03 - [Cached](#) - [Similar](#) - [Print](#)
 - VisualRevenue | 2007 August**: The Web Analytics Association (Jason Burdy and co.) recently releases 26 Standard Definitions to Promote Consistency among us Vendors (PDF) - this is GREAT ... visualrevenue.com/blog/2007/08 - [Cached](#) - [Similar](#) - [Print](#)
 - VisualRevenue | 2007 May**: First, if you do not read Robbins conversion blog - add it to your list, it belongs there. Secondly, there was a great post the other day: Conversion on a ... visualrevenue.com/blog/2007/05 - [Cached](#) - [Similar](#) - [Print](#)
 - VisualRevenue | 2007 April**: Finishing up and closing down after some successful days at ad:tech San Francisco - and with this event being one of the biggest shows in the interactive ... visualrevenue.com/blog/2007/04 - [Cached](#) - [Similar](#) - [Print](#)
 - VisualRevenue | 2007 October**: This is the 1st post in a series of posts about the factual (quantitative researched) state of the web analytics Industry. I also recommend reading the WAA ... visualrevenue.com/blog/2007/10 - [Cached](#) - [Similar](#) - [Print](#)

FIGURE 9-15. Indexing data on a website section

Both of these tools are very useful, and you should consider using them on every site you touch. The tools provide lots of valuable data, at no cost, and with little effort to boot. For example, [Figure 9-17](#) shows the opening screen for Bing Webmaster Tools (Bing WMT) for [Stonetemple.com](#).

As per [Figure 9-17](#), it looks like Bing has 2,570 pages from the site indexed. If you enter the `site:stonetemple.com` command in Bing itself, it shows something rather different (see [Figure 9-18](#)).

This search shows 11,800 indexed pages. Of the two, you should consider the report from Bing Webmaster Tools to have more accurate data than the results from a site search at Bing.

Link Building, Link Tracking, and Link-Based Metrics (Including Anchor Text Analysis)

Measurement of link-building campaigns is another must for the expert SEO practitioner. This can help you refine your campaigns to bring better results, compare link growth to your search rankings, and perform analysis to see how your campaigns are faring compared to the efforts of your competitors.

Search-engine-supplied tools

There are three basic tools for checking backlinks:

- Google Webmaster Tools is a powerful start. With Google Webmaster Tools, publishers can easily download a spreadsheet of all the links Google has in its database (note that Google might not include some links that it does not consider significant). With this tool

The screenshot shows a Google search results page with the following details:

- Search Query:** site:1up.com inurl:xbox360
- Results Count:** Results 1 - 10 of about 8,840
- First Result:** Search Results for Xbox360 from GameVideos. It includes a snippet of the page content: "Project Gotham Racing 3. Added on: 01/10/06 | Views: 2418. The popular series of PGR3 for the Xbox360. Gutterhero 360. Added on: 01/17/07 | Views: 875 ... gamevideos.1up.com/do/search?searchQuery=Xbox360 - Cached - Similar - ⌂ ⌂ ⌂
- Second Result:** Search Results for Xbox 360 Xbox360 CaseMod from GameVideos. It includes a snippet: "Search Results for Xbox 360 Xbox360 CaseMod from GameVideos. gamevideos.1up.com/.../search?...Xbox%20360%20CaseMod - Cached - Similar - ⌂ ⌂ ⌂
- Third Result:** Search Results for gt44 xbox360 game machinima from GameVideos. It includes a snippet: "Search Results for gt44 xbox360 game machinima from GameVideos. gamevideos.1up.com/.../search?...gt44%20%20xbox360%20game%20machinima... - Cached - Similar - ⌂ ⌂ ⌂
- Fourth Result:** Search Results for Pac-Man PacMan Xbox Xbox360 from GameVideos. It includes a snippet: "Search Results for Pac-Man PacMan Xbox Xbox360 from GameVideos. gamevideos.1up.com/.../search?...Man%20PacMan%20Xbox%20Xbox360... - Cached - Similar - ⌂ ⌂ ⌂
- Fifth Result:** Search Results for PS3 Xbox360 funny MyMIX from GameVideos. It includes a snippet: "Search Results for PS3 Xbox360 funny MyMIX from GameVideos. gamevideos.1up.com/.../search?...PS3%20Xbox360%20funny%20MyMIX... - Cached - Similar - ⌂ ⌂ ⌂
- Sixth Result:** Search Results for Gears War Marcus Fenix Xbox360 from GameVideos. It includes a snippet: "Search Results for Gears War Marcus Fenix Xbox360 from GameVideos. gamevideos.1up.com/.../search?...Gears%20War%20Marcus%20Fenix%20Xbox360... - Cached - Similar - ⌂ ⌂ ⌂
- Seventh Result:** Search Results for xbox360 from GameVideos. It includes a snippet: "Search Results for xbox360 from GameVideos. 1UP Network: 1UP | GameVideos | MyCheats | GameTab. GameVideos. JOIN NOW Free Membership | Login. My 1UP Page Join Clubs Message Boards Find Gamers ... gamevideos.1up.com/do/search?searchQuery=%20xbox360 - Cached - Similar - ⌂ ⌂ ⌂
- Eighth Result:** Re: ATTN: XBOX360 Board: MAKE YOUR PEACE - Xbox 360 - 1UP Message ... It includes a snippet: "Re: ATTN: XBOX360 Board: MAKE YOUR PEACE - Xbox 360 - 1UP Message ... 1UP NETWORK: 1UP | EGM | GFW | Filefront | GameVideos | MyCheats | GameTab. GAMES: #A B C D E F G H I J K L M N O P Q R S T U V W X Y Z | 1up.com ... boards.1up.com/thread/message?board.id=Xbox360...id... - Similar - ⌂ ⌂ ⌂
- Ninth Result:** Re: ATTN: XBOX360 Board: MAKE YOUR PEACE - Xbox 360 - 1UP Message ... It includes a snippet: "Re: ATTN: XBOX360 Board: MAKE YOUR PEACE - Xbox 360 - 1UP Message ... 1UP NETWORK: 1UP | EGM | Filefront | GameVideos | MyCheats | GameTab. GAMES: #A B C D E F G H I J K L M N O P Q R S T U V W X Y Z | 1up.com ... boards.1up.com/thread/message?board.id=Xbox360...id... - Similar - ⌂ ⌂ ⌂
- Tenth Result:** Search Results for Xbox360 from GameVideos. It includes a snippet: "Search Results for Xbox360 from GameVideos. 1UP Network: 1UP | EGM | Filefront | GameVideos | MyCheats | GameTab. GameVideos. JOIN NOW Free Membership | Login. My 1UP Page Join Clubs Message Boards ... gamevideos.1up.com/do/search?searchQuery=%20Xbox360...1 - Similar - ⌂ ⌂ ⌂

FIGURE 9-16. Site command combined with inurl

publishers can see only the links to their own site. Figure 9-19 shows what the spreadsheet looks like.

- Bing Webmaster Tools is also a great asset. It offers a similar capability for downloading a spreadsheet of the links that Bing has in its database for a site. Once again, publishers can use this tool only to see the links to their own site. Figure 9-20 shows what the report looks like.

Bing Webmaster Tools is limited in its export, however. Publishers can extract only up to 1,000 of their backlinks due to a limitation in the Bing API (as of July 2009). Bing does provide some cool filtering capabilities to help publishers work around that limitation, however.

The screenshot shows the Bing Webmaster Center interface for the website www.stonetemple.com. The left sidebar includes links for Site Bar, Related tools, Robots.txt validator, HTTP verifier, Keyword research tool, Webmaster resources, Forum, Blog, and Help. The main content area has tabs for Summary, Profile, Crawl Issues, Backlinks, Outbound Links, Keywords, and Sitemaps. The Summary tab is selected. It displays Site status information: Last crawled: 7/10/2009, Domain score: 5, Indexed pages: 2,670 (circled in red), and Blocked: No. Below this is a section titled 'Top 5 pages' which lists five web pages with their respective page scores, languages, regions, last crawl dates, and block status.

Web page	Page score	Language	Region	Last crawled	Blocked
Strategic SEO Planning, Technical SEO and Link Building	5	en	us	7/16/2009	No
Ramblings About SEO » Blog Archive » 11 Ways to Recognize a Bad ...	5	en	us	7/16/2009	No
Ramblings About SEO » Blog Archive » Google Local Business Center ...	5	en	us	7/14/2009	No
Ramblings About SEO » Blog Archive » 11 Ways to Recognize a Bad ...	5	en	us	7/16/2009	No
Ramblings About SEO » Blog Archive » Bounce Rate as a Ranking ...	5	en	us	7/16/2009	No

FIGURE 9-17. Bing Webmaster Tools

The screenshot shows a Bing search results page for the query "site:stonetemple.com". The results are labeled "ALL RESULTS". The first result is a link to "Strategic SEO Planning, Technical SEO and Link Building" from www.stonetemple.com, with a snippet about search engine optimization services. The second result is a link to "Ramblings About SEO » Blog Archive » 11 Ways to Recognize a Bad ...", with a snippet about recognizing bad SEO firms. The third result is a link to "Ramblings About SEO » Blog Archive » Latest Interview: InfoUSA ...", with a snippet about an interview with Pankaj Mathur of InfoUSA. The fourth result is a link to "Ramblings About SEO » Blog Archive » Google Local Business Center ...", with a snippet about Google's Local Business Center feature.

FIGURE 9-18. Bing site command results

- Yahoo! Site Explorer allows publishers to see all of their backlinks. However, only 1,000 of these can be exported into a spreadsheet. Seeing more than 1,000 requires you to step through the web results pages showing 100 linking pages at a time. Site Explorer will also allow publishers to see the backlinks for any website, not just their own. This is particularly interesting in evaluating the competition. [Figure 9-21](#) shows what the backlinks report looks like.

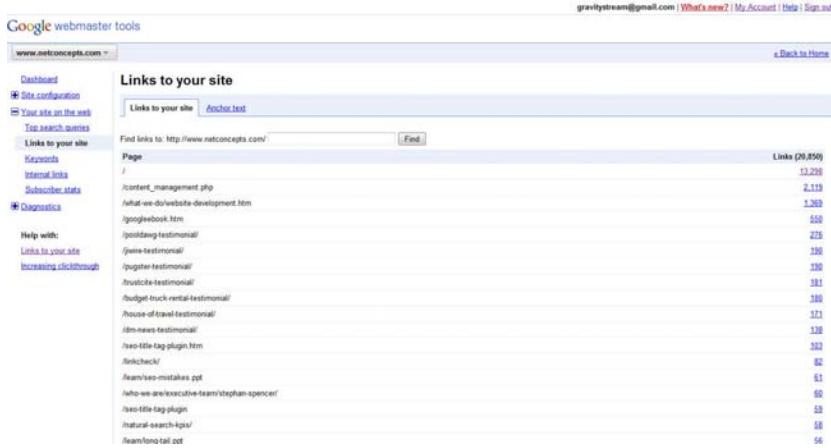


FIGURE 9-19. Google report on external links

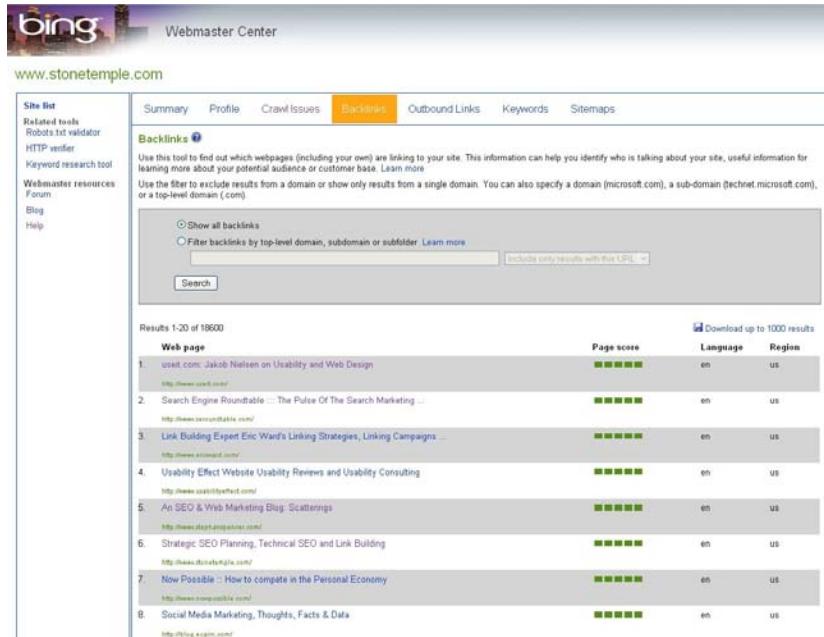


FIGURE 9-20. Bing report on external links

For quick and dirty link totals, it is handy to use a Firefox plug-in known as [SearchStatus](#). This provides basic link data on the fly with just a couple of mouse clicks. Figure 9-22 shows the menu you'll see with regard to backlinks.

The screenshot shows the Yahoo! Site Explorer interface. In the top left, it says "YAHOO! Site Explorer". The URL "http://whitehouse.gov" is entered in the search bar, and there is a "Explore URL" button. On the left sidebar, there are links for "Add to MySites", "My Sites", "Preferences", "Blog", "Badge", "Web Service API", and "Feedback". The main area is titled "Results" and contains a tip: "Tip: Your 'linkdomain:' query has been redirected to Site Explorer. You can continue to use normal web search for other queries." Below this, there are two buttons: "Pages (86,750)" and "Inlinks (4,984,538)". A dropdown menu for "Show Inlinks" is open, showing "From All Pages" and "To: Entire Site". There is also a "Submit webpage or Site Feed | Export first 1000 results to TSV" link. The results list shows seven items, each with a preview, a "Explore" button, and a link to the page's text/html version:

1. The White House
text/html [http://www.whitehouse.gov] - 50k
2. USA.gov: The U.S. Government's Official Web Portal
text/html [http://www.usa.gov] - 24k - cache
3. Homes & Communities - Department of Housing and Urban Development (HUD)
text/html [http://www.hud.gov] - 81k - cache
4. U.S. Department of Labor: Occupational Safety and Health Administration ...
text/html [http://www.osha.gov] - 47k - cache
5. Federal Bureau of Investigation (FBI)
text/html [http://www.fbi.gov] - 29k - cache
6. U.S. Department of Education
text/html [http://www.ed.gov] - 25k - cache
7. United States Department of Agriculture (USDA)
text/html [http://www.usda.gov] - 58k - cache

FIGURE 9-21. Site Explorer report on backlinks

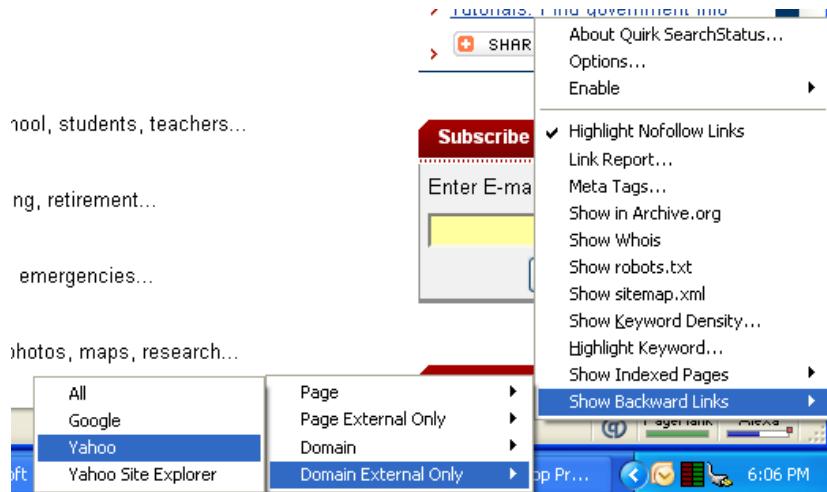


FIGURE 9-22. SearchStatus plug-in

Notice also in Figure 9-22 that the SearchStatus plug-in offers an option for highlighting NoFollow links, as well as many other capabilities. It is a great tool to help pull numbers such as these much more quickly than would otherwise be possible.

Third-party link-measuring tools

Advanced third-party tools exist for gathering link data. Here is a look at some of the most well-known ones.

Linkscape. [Linkscape](#) was developed based on crawl data obtained by SEOMoz, plus a variety of parties engaged by SEOMoz. You can see a complete list of Linkscape's sources at <http://www.seomoz.org/linkscape/help/sources>.

Linkscape is designed to be the tool for SEO practitioners to use when mapping links across the Web. It gets around the 1,000-URL limit imposed by Yahoo! and lets you get as many links as it can find and then extract them all into a spreadsheet.

[Figure 9-23](#) shows a sample report.

Included with this report are mozRank and mozTrust scores for each URL listed, and each domain. The tool also flags redirects so that you can see them. In addition, you can easily obtain additional data, such as the Google PageRank scores for the linking page and its domain.

The beauty of this tool is that it allows SEO practitioners and publishers to collect competitive intelligence on other people's domains, much like they do with Yahoo! Site Explorer today, but escape the 1,000-result limit and automatically get Google PageRank values as well as SEOMoz's proprietary mozRank and mozTrust scores.

We cannot underestimate the importance of mozRank and mozTrust. The PageRank score Google publishes through its toolbar is known not to be a reasonable approximation of the value of a given page, and is not the same as the actual PageRank Google uses in its ranking algorithms. The Google algorithm has simply evolved far beyond the original simplistic concept of PageRank that Google used in its early days.

Also of great value is *Domain mozRank* (DmR). This is the concept of assessing the overall trust level of a domain instead of a specific page. PageRank values published on the Google Toolbar are a measure for a particular page, not the domain, and there is no way to get that data from Google.

DmR uses the same mozRank algorithm but applies it to the "domain-level link graph." Doing so offers additional insight about the general authority of the domain. Just as pages can endorse other pages, a link that crosses domain boundaries (e.g., from some page on <http://searchengineland.com> to a page on <http://www.seomoz.org>) can be seen as an endorsement by one domain for another.

Many other factors are taken into account, such as trust level, relevance, and other types of references across the Web. MozRank and mozTrust attempt to incorporate these parameters and provide SEOs and publishers with a truly relevant measure of the importance of a page. In addition, all the data is extractable in a spreadsheet to allow for easy offline processing.

Link Diagnosis. Link Diagnosis can show you important information about your competitors' links, including page rank, anchor text, NoFollow links, and more. [Figure 9-24](#) shows an example of the basic Link Diagnosis output for one site.

The screenshot shows the Linkscape interface for analyzing the URL mit.edu. At the top, there's a banner stating "PRO Membership access granted". Below it, the Linkscape logo is displayed with the word "BETA" underneath. A sub-header says "Linkscape provides link data never available before on over 54+ billion URLs across 230+ million domains." A search bar contains the URL <http://mit.edu>, and a button labeled "Run Basic Report" is next to it. There's also a "+ Add URL" button.

The main title of the report is "Advanced Link Intelligence Report" for mit.edu, generated today at 15:10. The report includes tabs for "Dashboard", "Links to URL (2943)", "Links to Domain (2989)", "URL Anchor Text", "Domain Anchor Text", and "Custom Reports".

URL Data for mit.edu (most recently updated 2+ months ago)

- mozRank (mR): 5.64 OUT OF 10**
 - 84.53% from external links
 - 15.47% from internal links
- mozTrust (mT): 6.73 OUT OF 10**
- External Links: 47,850 links** from 1,120 subdomains
- Internal Links: 107 links**

Domain Data for mit.edu (most recently updated 2+ months ago)

- Domain mozRank: 6.34 OUT OF 10**
- Domain mozTrust: 6.84 OUT OF 10**
- Domain Juice: 5.85 EXTERNAL** (65.07% TOTAL DU OF 6.07)
- External Links: 62,665 links** from 3,985 domains

Top 5 Links to the URL mit.edu (ordered by mozRank passed)

Link URL	Title	Anchor Text	mozRank	mozTrust	Passing
http://iap.csail.mit.edu/	CSAIL Industry Affiliate Program Industry Affiliates Program		6.06	7.17	Passing: 36.98% of mR
http://www.w3.org	About Shawn	Anchor Text MIT	5.57	6.91	Passing: 3.73% of mR
http://www.sh1fty.com	Sh1fty's web site...	Anchor Text MIT	4.80	4.16	Passing: 1.13% of mR
http://ocw.cupide.org/OcwWeb/Global/privacy.htm	DATA NOT AVAILABLE		4.57	3.81	Passing: 1.07% of mR

FIGURE 9-23. Linkscape report

You can download each subsection of this report into a CSV file, for later examination in Excel. Pulling the basic backlinks report (shown at the top left of Figure 9-24) into a spreadsheet provides data that looks like Table 9-1.

The screenshot shows the Link Diagnosis interface with the following sections:

- Top Navigation:** Home, Download, FAQ, Blog, Report a Bug.
- Logo:** linkdiagnosis.com
- Input Fields:** Website: http://stonetemple.com, Start, e.g. http://www.link-diagnosis.com
- Report Summary:** Total backlinks: 1833, Total unique anchor texts: 477.
- Backlinks Table:**

url	PR	anchor	linktype	links	strength	(CSV)
http://www.dimensionoffilms.com/	7	N/A	missing	564	4	
http://www.diminu.edu/tss/support...	6	N/A	missing	508	2	
http://www.access-masters.com/...	6	Interview with Go...	good	137	10	
http://www.access-masters.com/...	6	Eric Enge Interview...	nofollow	9	5	
http://en.wikipedia.org/wiki/Web...	6	Matt Cutts Interview...	nofollow	92	15	
http://www.torprankblog.com/search...	6	Ramblings About SEO	good	543	3	
http://www.customsearchguide.com/...	6	How do I create a...	good	32	43	
http://www.vcorpblog.com/2009/02/26/...	6	Eric Enge	nofollow	198	7	
http://www.searchenginemashdown.com...	5	Interview today w...	good	160	5	
http://www.liauhill.net/explainitm...	5	Competitive Intel...	good	562	2	
- Top Pages Table:**

url	count	strength	(CSV)
http://www.stonetemple.com/	499	1744	
http://www.stonetemple.com/266	95		
http://www.stonetemple.com/192	35		
http://www.stonetemple.com/265	25		
http://www.stonetemple.com/123	30		
http://www.stonetemple.com/97	22		
http://www.stonetemple.com/32	33		
http://www.stonetemple.com/27	20		
http://www.stonetemple.com/24	5		
http://www.stonetemple.com/73	9		
- Most Popular Anchors Table:**

anchor	count	(CSV)
Stone Temple Consulting	119	
Eric Enge	118	
Comments	34	
2007 Web Analytics Shootout	31	
Ramblings About SEO	27	
Stone Temple	25	
Interview...	19	
http://www.stonetemple.com/articles/interview...	15	
StoneTemple Consulting	15	
2007 Web Analytics Shootout - Final Report	14	

FIGURE 9-24. Link Diagnosis report**TABLE 9-1.** Link Diagnosis link report export

url	pagerank	anchor	linktype	Obl
http://searche	7	Null	0	288
http://blog.se	6	Google's S	1	906
http://www.c	6	How do I	1	31
http://www.s	6	Eric Enge	1	171
http://blog.se	6	spoken w	1	1060
http://en.wik	6	Eric Enge	2	265
http://blog.se	6	StoneTem	1	412
http://stonete	6	web mark	1	176
http://glinder	5	Posted	1	324

url	pagerank	anchor	linktype	Obl
http://blog.se	5	Andrew T	1	99
http://www.n	5	Stone Tem	1	181
http://kn.thei	5	Null	0	86
http://www.g	5	http://ww	2	33
http://stonete	5	Ramblings	1	210

Other third-party link-building tools

A variety of additional third-party link-building tools are also available. Enquisite Optimizer (<http://www.enquisite.com/products/optimizer/>), for one, offers a variety of link-building reports. For example, one of the best ways to discover when you have received a new link is to monitor new referrers. Enquisite Optimizer builds this in, and will report to you links that never sent you traffic in the past. To help evaluate their importance, the tool provides you with the Alexa Rank, PageRank, mozRank, and Compete Rank for each link. Enquisite Optimizer offers many other link-building reports, including those that enable you to:

- See places on the Web where you are referenced, but the reference is not implemented as a link
- See who is linking to your competitors but not to you
- See who is linking to sites in your neighborhood but not to you

Developed by Bruce Clay, Inc., [LinkMaps](#) allows you to map the backlinks of any website, including your competitors'. LinkMaps gathers the initial data from the search engines and then adds some advanced filtering, including:

- Removing pages that return 404 errors
- Removing pages that do not have a link
- Limiting results to no more than four pages per domain
- Filtering out guestbooks
- Identifying possible link farms

LinkMaps also shows you which search engines found the link and which did not. In addition, LinkMaps will build pages that contain links to the pages that link to you. The page is **NoIndexed**, so it will not show up in the search engines, but you can use it to help the search engines discover the links to your site.

Not too long ago, the European link-building agency [Cemper.com](#) launched its link building toolkit [LinkResearchTools.com](#), which includes a link juice analyzer, common inbound links (between you and a competitor), common outbound links, and an anchor text analyzer.

Squid is a tool developed by [Searchreturn](#) to help link builders project manage the link-building process. You let Squid know the site you are building links for, and tell it to fetch a Yahoo! link report for that site. Then you give it a list of your competitors' sites (or of sites that you want to backlink for some reason).

Squid then builds a list of people who link to those other sites but do not link to you. The initial data looks something like [Figure 9-25](#).

The screenshot shows the Squid interface with the following components:

- Search Options:** Includes fields for "Search for Link Targets that Include:" (set to "Don't care") and "And have a Status of:" (set to "Don't care").
- Link Targets:** A list of websites under "And are linked to:" which includes 1up.com, games.com, games.net, gamespot.com, xbox-scene.com, gamespy.com, ign.com, metacritic.com/games/, www.gamezone.com, www.videogamerreview.com, www.gamedaily.com, and www.tomshardware.com.
- Filter Results:** Buttons for "Save Link Targets", "Previous 1 2 3 4 5 6 7 8 9 ... 144 145 Next", "Show all results on one page", and "Move all Raw to Investigate".
- Results Table:** A table with columns: Edit, URL, Common Links, and Status. The table contains 5 rows of data:

	Edit	URL	Common Links	Status
1	Edit	http://en.wikipedia.org/wiki/Metal_Gear_Solid_3_Snake_Eater	4	— Inv F L? LI X
2	Edit	http://www.gamepress.com/	4	○ ⊕ ○ ○ ○ ○ ○
3	Edit	http://www.info-s.com/	4	○ ○ ○ ○ ○ ○ ○
4	Edit	http://your-free-games.com/	4	○ ⊕ ○ ○ ○ ○ ○
5	Edit	http://en.wikipedia.org/wiki/Square_Enix	3	○ ⊕ ○ ○ ○ ○ ○

FIGURE 9-25. Squid screen shot

The results are sorted in descending order based on common links. The number in the Common Links column represents the number of other sites that the page links to. So, when someone links to five competitive sites but not to you, that makes for a juicy link target!

The columns on the right of [Figure 9-25](#) contain the status of the link request process, using the following definitions:

- Inv = Investigate
- F = Follow up
- L? = Make a link request
- LI = Got the link
- X = Rejected, either because they said no or you decided they were not a good target

Another impressive toolset for link building is [Raven](#). One of Raven's key features is the Link Manager, which tracks links that you plan on acquiring, have requested, or have successfully

built. Raven includes conversion and ROI tracking, as well a Firefox toolbar that makes it easier to add websites to your queue of link acquisition targets. Raven, like Squid, will go out and automatically check the identified pages to see whether you have acquired a link on them, and it will tell you the results—thus automating the process of identifying targets and tracking the progress of one's campaign on an ongoing basis.

Google Blog Search. It is well known that the `link:` command works poorly in Google Web Search. For whatever reason, Google has decided to return only a small sampling of data when people use that command. However, interestingly enough, you can get more data on your backlinks using Google Blog Search (and it reports more than just the blog links); see [Figure 9-26](#).

The screenshot shows the Google Blog Search interface. At the top, there is a search bar with the query "link:stonetemple.com". Below the search bar are buttons for "Search Blogs" and "Search the Web", and links for "Advanced Blog Search Preferences". The results section is titled "Blog results" and shows "Results 1 - 10 of 2,546". On the left, there is a sidebar with filters for "Published" (Last hour, Last 12 hours, Last day, Past week, Past month, Choose Dates) and a "Subscribe" section with links for "Blog Alerts" and "Atom | RSS". The main content area displays three search results:

- Interview with Peter Linsley from Google** (14 Jul 2009) - This means that Google Image Search has about 8% of the traffic of Google Web Search. For people with image rich sites this means there is a substantial opportunity. Bear in mind that the number of people competing for that image search ... [searchnewz: Latest Search News RSS Feed - http://www.searchnewz.com/](#)
- Google Loses Engineering Director Who Once Caused Steve Ballmer To ...** (14 Jul 2009) - from <http://www.techcrunch.com/2009/07/13/google-loses-engineering-director-who-once-caused-steve-ballmer-to-melt-down/>. Posted: 13 Jul 2009 10:44 PM PDT. After nearly 5 years with the company, Engineering Director Mark Lucovsky is ... [newsnode's posterous - http://newsnode.posterous.com/](#)
- Google Loses Engineering Director Who Once Caused Steve Ballmer To ...** (13 Jul 2009 by JW) - After nearly 5 years with the company, Engineering Director Mark Lucovsky is leaving Google for a role with VMware, we've learned. Lucovsky has been an integral part of Google's APIs, including the all-important Search APIs. ... [Trinitude Network - http://www.trinitude.com/web/](#)

FIGURE 9-26. Google Blog Search screen shot

Technorati. Technorati provides a look at the number of blog reactions to a website. For a site with a major blog such as SEOMoz, this can provide extremely valuable data. [Figure 9-27](#) shows an example.

Exalead. Exalead, a European search engine, offers capabilities that include a `link:` command. What make Exalead particularly interesting are the advanced filtering capabilities it provides, as you can see in [Figure 9-28](#).

Note the ability to pick language as well as all the other options, which could be useful for segmenting link data.

Measuring the value of a link

One of the big questions that people ask is what is the value of a particular inbound link? There is no simple way to answer that question, but you can look at some of the metrics that can give you a feeling for how important a link might be:

The screenshot shows the Technorati homepage with a search bar at the top containing "search the blogosphere...". Below the search bar is a navigation menu with links like "Channels", "Blogs", "Photos", "Videos", "Favorites", "Popular", "Blogger Central", "Support", and "Tools". A search result for "seomoz.org" is displayed, showing 18,762 blog reactions. The results are filtered by "entire post", "any authority", and "all languages". The first result is "High Impact/Low Effort SEO from a Developer's Perspective" by Wulf's Bookmarks, posted some time ago, with no authority yet. The second result is "Internet Marketing, Strategy & Technology Links – July 22, 2009" by Sazbean, posted some time ago, with Authority: 16. The third result is "CAPTCHAs' Effect on Conversion Rates" by BBT Webmasters, posted some time ago, with Authority: 1.

FIGURE 9-27. Technorati blog reactions data

Most important elements to determining a link's value

Where does the linking page rank for the term/phrase you want to rank for?

If the page is ranking #1 at Google for *sliced bread* and you want to be #1 at Google for *sliced bread*, guess what? That's the #1 most valuable link you can get. Keep going down the list to about positions 25 to 30 and you're still getting solid gold in link value.

Where does the linking page rank for one to two important, competitive terms in its title tag?

This will give you a very solid idea about how much overall link juice and respect the search engines are giving the page. It is also a good way to identify the global link value that could be provided.

Where does content on the linking domain generally rank for competitive terms in its pages' respective title tags?

As in the preceding list item, we're trying to identify how positively the engines view pages on the domain. If the pages generally rank in the top 20 results, you can rest assured that search engines think the domain's value is pretty high, and that links from that domain will pass significant value.

Does the linking site carry any brokered sets of links?

Sites that sell links may lose their ability to pass link juice. This really applies to any type of low-quality, manipulative linking. If you can see it, chances are that Google might see it someday too. In addition, Google may penalize a site retroactively for repeated bad behavior, even if that behavior has been stopped.

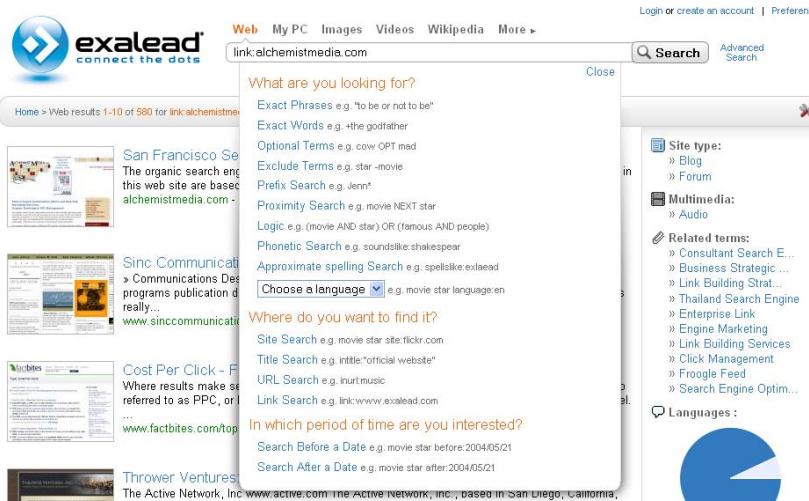


FIGURE 9-28. Exalead link data

What is the relevance of the linking page/site to your target page?

Answering this question requires you to think critically about the visitors to both the potential linking page and the domain. If the relevance of the subject matter to your site is high, the link will provide more semantic and topic-specific value.

Elements of secondary value for links

Links to high-ranking competitors

Although this isn't always an indication of direct value, it can be a good signal. Your competitors are obviously ranking based on the strength of their links, so researching those sources can provide insight into where they derive that value.

Page strength

SEOMoz offers a free tool that computes a metric called **Page Strength**. Although it is tough to make too much out of the numerical calculation the tool offers, the other data returned is convenient and certainly valuable. If there are a lot of links from Wikipedia and DMOZ and the site has high PageRank, lots of inbound links, and blog links, there's clearly some value to getting a link. **Figure 9-29** shows a screen shot of the output of the tool.

PageRank of the domain

There is a notion that the PageRank of a domain would be the overall PageRank value Google would associate with the domain. This is not a value you can get at directly, but you can approximate it by looking at the PageRank of the website's home page. Look at this to make sure it is not penalized and to see the overall link juice. You can

also get an assessment of the link strength of a domain using DmR from SEOmoz's Linkscape tool, which we discussed earlier in this chapter.

A PageRank 6 (or higher) home page clearly has some link love and respect, a Page Rank 2 obviously has less, a gray bar can be a good red flag, and seeing a PageRank 0 tells you that the domain is either new or completely invisible. However, as link-building expert Eric Ward points out, avoiding lower PageRank domains simply because they are PageRank 3 or 4 is not necessarily a good thing, as you could be missing out on very relevant links that, in volume, contribute to your relevance.

Inlinks to page (via Yahoo!)

Next, look at the links to the specific page you want to get a link from (or perhaps that already links to you). You want to know whether the domain links into this individual page heavily, or whether it is practically an orphan. See whether it is a page that other sites reference, both of which can help illuminate potential value.

Inlinks to domain (via Technorati or Google Blog Search)

The Technorati and Google Blog Search link data can show trends—if a site has been picking up many new links over the past few months, it might be a much better candidate than the PageRank or other link data might indicate.

Inlinks to domain (via Yahoo!)

This is a pretty indirect measurement, but it's not completely useless. However, since the number often takes into account lots of links from a single domain, it can be misleading.

PageRank of page

Since pages that are relatively new (three to four months old) are shown on the Google Toolbar as having a PageRank of 0 (at least until Google updates the PageRank it shows on the Google Toolbar), and since so many valuable pages may have PageRanks that are only in the 1–3 range, it seems unwise to get caught up in the PageRank of a specific page. It is better to look at the domain and the attention it gives your page. However, this can be valuable if the page has been around for a while.

Number of external links on the page

Pages generally pass some of their link juice to each page they link to. Therefore, if you have a link from a page with 10 outbound links, it is far better than having a link from an equivalent page with 100 outbound links, so this is a parameter worth knowing about.

Rankings

Many people get obsessed with checking their rankings. Rankings do have some correlation with traffic. However, do not spend too much time on this activity. For one thing, many tools that are available today for checking rankings scrape Google results, which is against Google's Webmaster Guidelines. Google results vary by user, and the scraped results do not represent

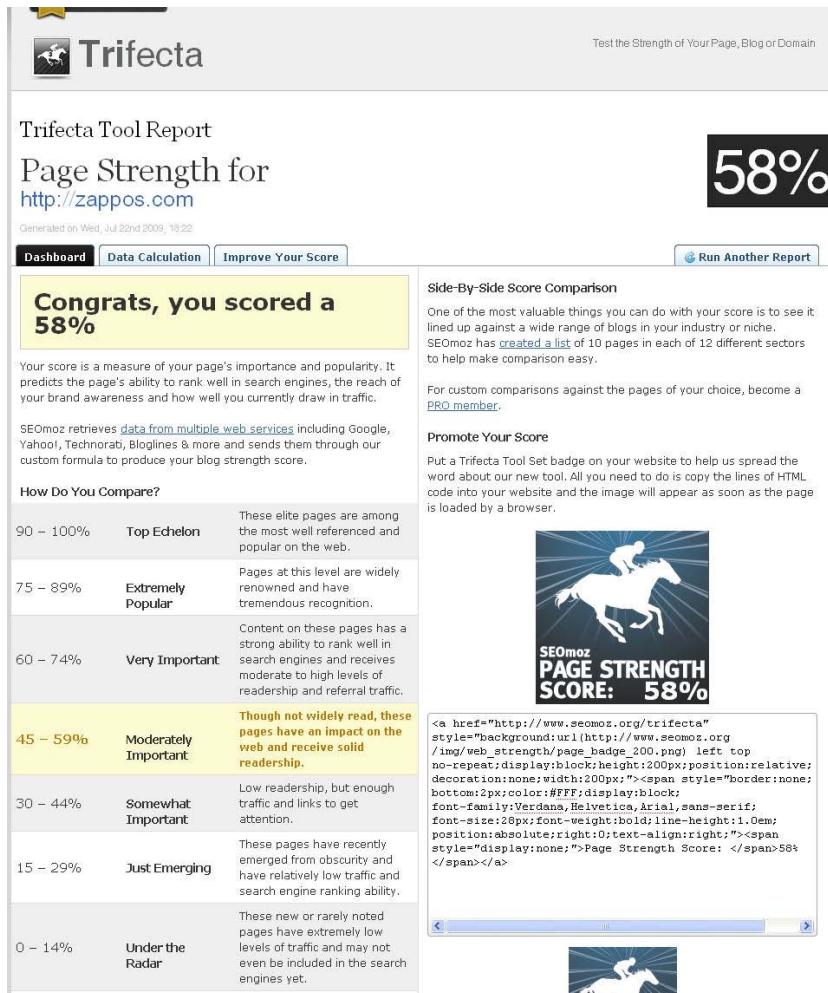


FIGURE 9-29. SEOmoz Page Strength tool

what everyone is seeing. Therefore, rankings data is not as useful an indicator as many of the other metrics we have discussed.

Rankings data can be helpful in a few scenarios, however:

- Where a #1 (or a very high) ranking for a trophy search term provides brand value
- When there is a sudden significant change (such as a move from position #5 to #50, or a similar change in the other direction)
- Where a long-term trend can be established, such as a steady improvement in position, or a steady decline

While evaluating this data, bear in mind that the search engines are constantly being tweaked. Google is believed to tweak and test its algorithms every day. This can result in a lot of movement in rankings on a regular basis, and this type of movement is usually meaningless from an SEO perspective.

What follows are some examples of tools that allow you to obtain rankings data. For one, you can get the data directly from Google Webmaster Tools (see Figure 9-30).

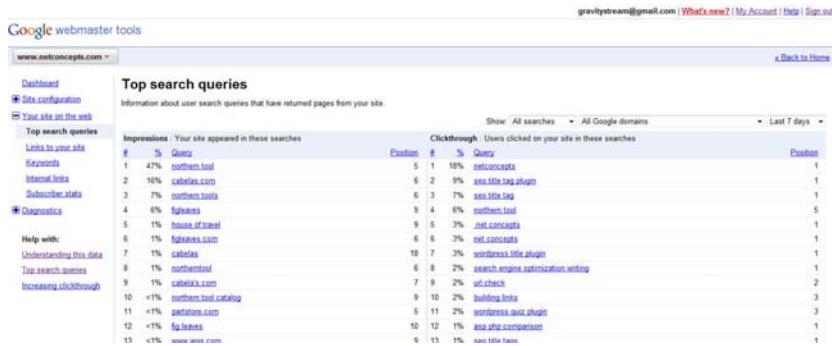


FIGURE 9-30. Top search queries report

Google handily shows the top 20 queries for which your site came up, as well as the queries that were clicked on the most. This is nice data to give you a snapshot of where things stand.

Enquisite Optimizer provides some visibility into rankings data, as shown in Figure 9-31.

Last but not least are the tools from [Thumbshots](#). For example, the Thumbshots rankings tool provides rankings data as well, but offers the novel additional feature of enabling you to pull the rankings for two sites at once, and get comparative data in a graphical form, as shown in Figure 9-32.

Many other tools allow you to collect rankings data. Some of the most important ones are:

- [AdGooroo](#)
- [Rank Checker](#)
- [Web CEO](#)

Whatever tool you choose to use, or even if you do your checks manually, remember to stay focused. Rankings move back and forth on a regular basis due to changes in search engine algorithms. Focus on changes of significance as we previously outlined, and use traffic and conversion measurements for day-to-day measurement of progress.

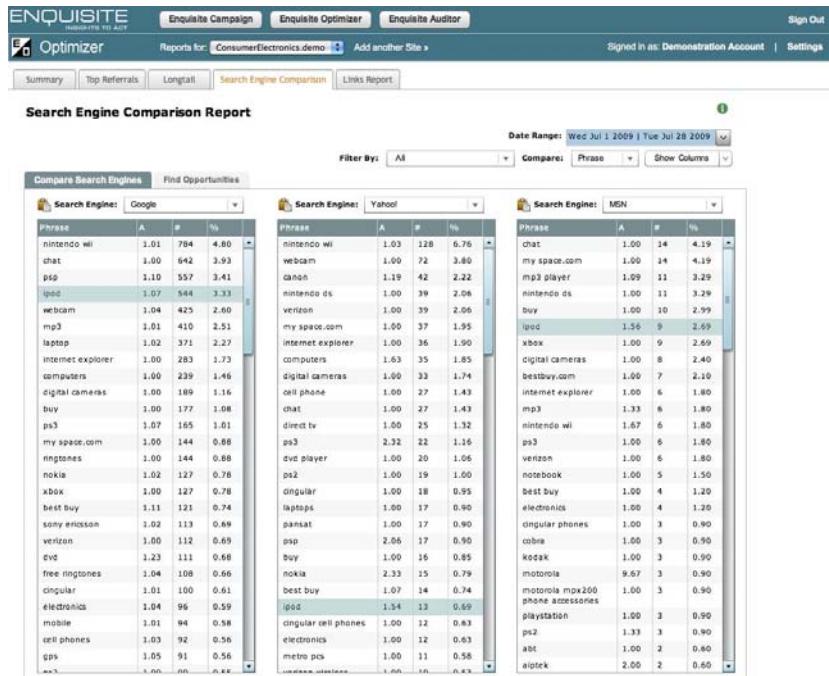


FIGURE 9-31. Enquisite search phrases report

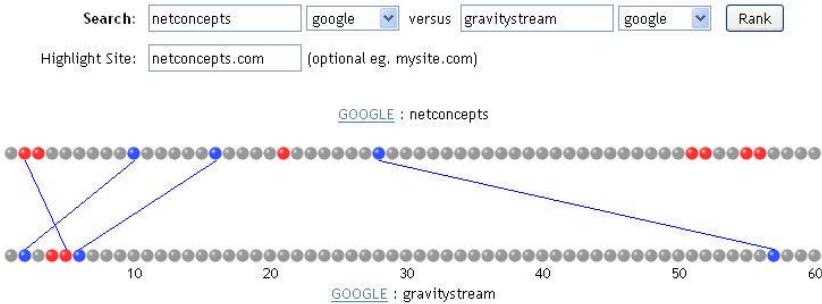


FIGURE 9-32. Thumbshots rankings report

Crawl Errors

Complex websites will ultimately develop errors. This will come from moving pages where the publisher forgets to redirect the old page to the new page; or the publisher makes a mistake in her *robots.txt* file and prevents the crawling of key parts of her site. It can also happen that

someone links to the publisher's site, but she implements the link incorrectly by linking to a page that does not exist (perhaps as a result of a typo).

Tools from the search engines

All of these things are easy to diagnose in either Google Webmaster Tools or Bing Webmaster Tools. Figure 9-33 shows the Crawl Issues page in Bing WMT.

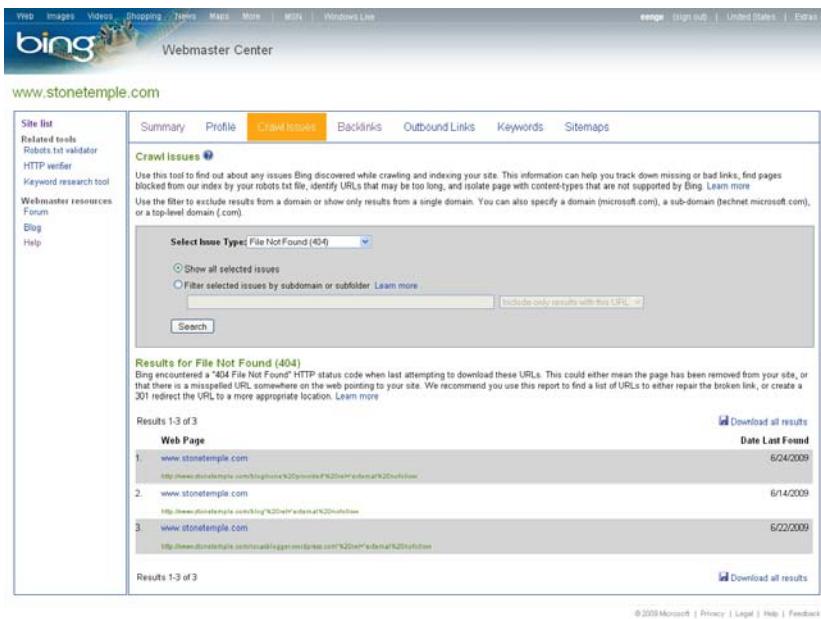


FIGURE 9-33. Bing Crawl Issues page

You can see that Bing will provide data on 404 pages, pages blocked by REP (which is short for Robots Exclusion Protocol), dynamic URLs that might be troublesome for Bing, malware-infected pages, and unsupported content types.

Google provides similar data, with some notable differences (see Figure 9-34).

Notice the additional types of errors provided. In addition, if you look at the 404 report in Google WMT, you have the option of seeing what pages link to the page generating the 404 error, as shown in Figure 9-35.

The great thing about the 404 reports in both tools is that you can spot pages that have been implemented with the wrong URL on your site, and you can spot whether someone links to one of your pages using the wrong URL. In the latter case, as we suggested in [Chapter 7](#), 301-redirect the page that received the bad link to the correct page, and pick up an easy link. In

The screenshot shows the 'Crawl errors' section of the Google Webmaster Tools interface. It displays a table with three columns: URL, Detail, and Detected. A note at the top states 'No known errors'.

URL	Detail	Detected
No known errors		

FIGURE 9-34. Google crawl errors

The screenshot shows the 'Crawl errors' section of the Google Webmaster Tools interface. It displays a table with two entries:

URL	Discovery Date
http://tech.groups.yahoo.com/group/webanalytics/message/11892	May 9, 2008
http://tech.groups.yahoo.com/group/webanalytics/message/12137	Jul 13, 2009

At the bottom right, there is a note: 'Updated Jul 17, 2009'.

FIGURE 9-35. Google 404 errors report

short, 404 errors provide great opportunities to request updated links from sites linking to old URLs.

Third-party tools to check for crawl errors

Some excellent third-party tools are available for seeing how a search engine crawler will look at your site. One of these is [SEO-Browser](#). Use this tool to see how search engine crawlers look at your web pages. This can be useful during a technical SEO evaluation of a site, or when debugging site indexation problems. [Figure 9-36](#) shows the basic view from SEO-Browser for the [TripAdvisor](#) home page.

Interestingly, this report indicates that the website shows that the surfer used an unknown browser. It might be interesting to determine why that is happening. If you look at the Amazon home page using the Advanced mode, you can see some interesting data as well (see [Figure 9-37](#)).

The Advanced view provides a wide range of stats above the view of the web page itself. Down toward the bottom of [Figure 9-37](#), you can see that Amazon offers a screen reader version of the page.

Another excellent tool is the SEOMoz Crawl Test Tool. It provides an extensive set of statistics on any website you choose to check. [Figure 9-38](#) shows a small sample of the data provided.

Analysis mode: **simple** [switch to advanced]
 Google PR is 7 out of 10
[SEO Browser Bookmarklet](#) ([What's This?](#))
[Googlebot 2.1 \(New version\)](#)

seo-browser.com™

See your website like a Search Engine sees it.

Address: <http://www.tripadvisor.com> Parse URL

copyright © 2005 - 2009 [Aordan Marketing](#) and Ian McAvity

For better analysis, switch to the [advanced mode](#).

[TripAdvisor - Hotel Reviews] Over 20 million traveler reviews & opinions! Sign in.
 Register Now! When you register you can: Share your stories
 Save your favorite places and plan your next trip
 Connect with other travelers
 Get the Member's gear

Register Now Already a member? Sign in
 [International Sites]

[Write a Review](#) [Hotels](#) [Flights](#) [Restaurants](#) [Trip Ideas](#) [Free Travel Guides](#)

Get advice from real travelers — more than 15,000,000 travelers from 190 countries planned trips here this week. Search

No-Script	11/08/04 22:00
---------------------------	--------------------------------

Browse by Destination [Canada](#)
[United States](#)
[Caribbean](#)
[Mexico](#)
[Central America](#)
[South America](#)
[Europe](#)
[Africa](#)
[Asia](#)
[Middle East](#)
[Africa](#)
[South Pacific](#)

No-Script	11/08/04 22:00
---------------------------	--------------------------------

No-Script	11/08/04 22:00
---------------------------	--------------------------------

upages/home/sidebar

Sponsored by: [Sheraton](#) Traveler Lists

Recent Traveler Lists [Zion National Park in less than 6 hours](#) Jul 2, 2009 [checkin](#)
[Island RYC Experiences](#) Jun 28, 2009 [checkin](#)
[\[57685/10971\]](#)

Plan the Perfect Trip

[Hotels](#)
[Flights](#)
[Restaurants](#)
[Things to Do](#)
[Cruises](#)
[Vacation rentals](#)
[Forums](#)

Find Hotels Travelers Trust test City: Check-in: mmddyy Check-out: mmddyy Adults: 1 2 3 4
 Traveling for Business? (NEW!) [Find top-rated hotels, restaurants & more...](#)

Find Flights From: Depart: mmddyy Anytime Early Morning Morning Afternoon Evening Night To: Return: mmddyy Anytime Early Morning Morning Afternoon Evening Night Travelers: 1 2 3 4 5 6 Prefer nonstop [MORE](#)
[OPTIONS](#) • (One-way, multi-city, etc) [\[Read summer excuses under 1199\]](#)

Routes & Flares

The best and easiest routes

[\[Reviews: 10+ yrs old\] 100m on Hawaiian Village Beach Resort & Spa](#) Honolulu, Oahu (5.0 of 5 stars)
 "The room was large and the view MAGNIFICENT!" [here](#)

[\[Reviews: 10+ yrs old\] 145 \[Reviews: down vs last\] Gulf House Hotel & Suites](#) Louisville, Kentucky (1.0 of 5 stars)
 "Smelly, Hot, Shabby & Dirty-Otherwise Ok." [here](#)

By [BruceLafon](#) Featured Traveler Photo [[Fort McHenry - July 21, 2007](#)] Fort McHenry National Monument: Baltimore Submitted for [matrix](#): [Find out more about Fort McHenry National Monument](#) Spotlight destination [Carefree Cave Creek] Carefree and Cave Creek, AZ Just 20 miles west but worlds away from Phoenix, Carefree and Cave Creek have something for everyone. From rustic ranches to luxurious spas, fancy shops to fine art galleries, picture-perfect resort towns to cool bars, and rollickin' saloons to five-star restaurants, it's all waiting for you here in one of Arizona's last frontiers. Description provided by [CarefreeCaveCreek.org](#) [Read more about Carefree](#) » Quick Links: [Share my experience](#) [Write a review](#) [Ask a question or leave a comment](#) [Write a Traveler Article](#) [Create a Traveler List](#)

FIGURE 9-36. SEO-Browser

The tool also provides a detailed look at some of the more important pages on the site, including the indexing status in the major search engines, the most important keywords, metadata, and the number of internal links to the page.



FIGURE 9-37. Amazon Advanced mode

Tools such as these that provide a crawler's-eye view of your web page or website can be effective in helping you to rapidly diagnose problems with your site.

Tracking the Blogosphere

The blogosphere is a very dynamic part of the Web—always fresh and constantly updated. News tends to break in the blogosphere, and in Twitter, before it hits mainstream media or the Web at large. Tracking the blogosphere will help you stay current on both your industry and the SEO industry. Major influencers tend to hang out in the blogosphere as well as have their own blogs. Identifying these influencers is the first step to reaching them.

The screenshot shows the SEOMoz Crawl Test tool interface. At the top, there's a navigation bar with links for Blog, Tools, Guides, YOUMOZ, Marketplace, Store, and About. A yellow ribbon banner on the left says "PRO Membership access granted". The main header is "CrawlTest" with a small icon. Below it, a sub-header says "Discover the problems search engine robots run into while crawling your website, the http response codes your pages return, cache dates, title tags and more." A "Crawl My Site" button is visible.

The main content area has a "Enter URL" field containing "http://www.staples.com" and a "Crawl My Site" button. Below this, a message says "You can view this report by visiting: http://www.seomoz.org/crawl-test/report/b4ac7452-4b03-445b-a10a-8863c4486540".

50 of 50 pages crawled

	Number of Internal Links (across all pages)	7155	Pages indexed in Google	31
Number of outbound links (across all pages)	543	Pages indexed in Yahoo!	32	
Pages with unique <title> tags	41	Pages indexed in Bing	42	
Pages with duplicate <title> tags	9	Average Google cache date	0	
Pages lacking <title> tags	0	200 (All's well)	48	
Pages with unique meta descriptions	20	300 (Redirection)	1	
Pages with duplicate meta descriptions	20	400 (Something's wrong)	0	
Pages lacking meta descriptions	0	500 (Server Error)	0	

1 **Page Title** Office Products - Office Supplies, Furniture, Technology & Electronics | Staples[®],
URL <http://www.staples.com>
HTTP Status Code 200
Indexed in Google YES
Indexed in Yahoo NO
Indexed in Bing YES

Meta Description Staples makes it easy to find the business essentials you need, like office supplies, ink & toner, copy paper, technology & electronics, office furniture. Get FREE delivery on most orders over \$50.
Meta Robots index,follow
Number of Internal Links 170
Number of Outbound Links 11
Last crawled on Thu Jul 23 05:50:42 -0700 2009

2 **Page Title** Office Products - Office Supplies, Furniture, Technology & Electronics | Staples[®],
URL <http://www.staples.com>
HTTP Status Code 200
Indexed in Google YES
Indexed in Yahoo NO
Indexed in Bing YES

Anchor Text Products Main
Meta Description Staples makes it easy to find the business essentials you need, like office supplies, ink & toner, copy paper, technology & electronics, office furniture. Get FREE delivery on most orders over \$50.
Meta Robots index,follow
Number of Internal Links 170
Number of Outbound Links 11
Last crawled on Thu Jul 23 05:50:42 -0700 2009

FIGURE 9-38. SEOMoz Crawl Test tool

Here are four major uses for tracking blogs:

Reputation monitoring

A significant component of reputation monitoring relates to the discussion in this section about domain and brand mentions across the Web. Reputation monitoring takes it one step further, where you look to identify problems and risks to your reputation as they materialize on the Web.

For example, knowing when someone starts to bad-mouth your brand somewhere on the Web is important. Tracking down those issues early and addressing them quickly is a wise thing to do.

Tracking buzz and public relations campaigns

A closely related activity is tracking buzz and the response to your PR campaigns. When you make a major press push of some sort or you succeed in getting a write-up about your website on an influential blog, you should monitor the ripple effect across the rest of the Web.

Using the type of brand tracking we've already discussed can be a great way to do that. For example, did the number of mentions you received on the Web this week increase significantly over the number of mentions the week before?

Identifying potential influencers

Identifying influencers is a key part of link building. As we've discussed throughout the book, establishing yourself as a recognized expert is an important part of your content strategy. Part of that is creating world-class content, but this won't mean much unless you get the word out somehow. A great way to do that is to reach out and develop relationships with the key influencers.

Competitive analysis

If you can do this type of research for your own site, why not do it for your competitors' sites? You can see reputation problems that are developing for them, or see the impact of their media campaigns as they roll them out.

Tracking Your Blog(s)

Blogs offer a lot of different benefits to the online marketer. They can offer a great way to position yourself or the team behind your site as experts. For that reason, you can reach major influencers as well as a broad audience of people at the same time. This can become a nice source of links as a part of your link-building campaign.

Blog subscribers

Measuring the number of subscribers to your blog is one basic way to monitor the blog's progress. One way to do this is to create your RSS feed with [FeedBurner](#) and have users use that RSS feed instead of any other solution, such as the one that comes with your blog software. By having users subscribe to the FeedBurner-based RSS feed, you can get much more granular statistics on your subscriber base, as shown in [Figure 9-39](#).

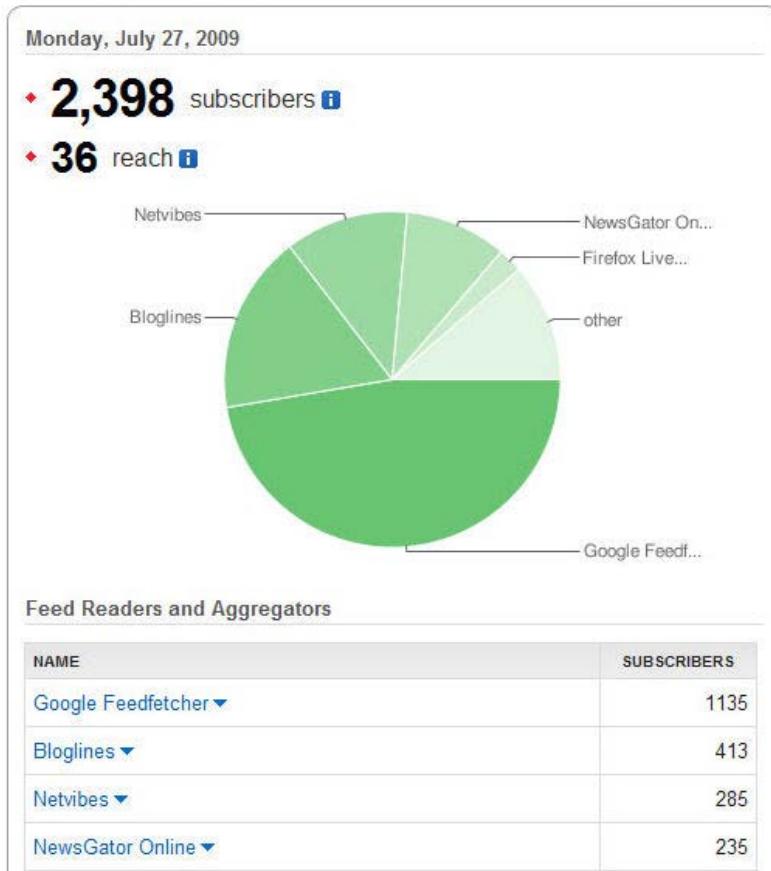


FIGURE 9-39. FeedBurner stats

As you can see in the figure, you get data on the growth of your readers over time, as well as a breakout of your subscribers showing which feed readers they are using. Be aware that although this measurement is pretty inaccurate, it can still give you an idea of where you stand. Note that in [Figure 9-39](#), Bloglines is used by about 10% of this blog's readers.

You can also get some data regarding how many of the people who are subscribing to your blog or to your competitors' blogs are using [Google Reader](#) or [Bloglines](#) to do so. [Figure 9-40](#) shows an example of how such a report looks in Bloglines.



FIGURE 9-40. Bloglines subscriber information

Figure 9-40 shows that 394 people are subscribing to Google Blogoscoped (<http://www.blogoscoped.com>) using Bloglines. You can perform a similar check in Google Reader to see how many people are subscribing to the blog using Google Reader. As there are many different types of readers out there besides Google Reader and Bloglines, this data is limited in its scope, but it is still extremely useful for telling you something about how many people subscribe to your blog, and for its use in competitive analysis.

N O T E

Another way to take this analysis deeper is to look for specific reader user agents in your logfiles. For example, you could use this to find out how many My Yahoo! readers you have.

Blog links

There are a few ways to extract data on the number of links to your posts. One is to use Technorati to see the number of Reactions (i.e., links) to your posts, as we outlined earlier in this chapter. You can also use either Google Webmaster Tools or Bing Webmaster Tools to track the number of links to your blog.

As we outlined in the discussion on link tracking, you can use these tools to get link data on your website. You can also take a more granular look at your data. For example, if you have a blog at <http://www.yourdomain.com/blog>, you can pull a report from Google Webmaster Tools (or your favorite tool of choice) and sort your spreadsheet on the URLs of the pages receiving the links. When you are done with this, your data might look something like Table 9-2.

TABLE 9-2. Sorted link report

8433	http://www.stonetemple.com/articles/video-	http://www.	7-Jul-09
8434	http://www.stonetemple.com/articles/video-	http://www.	12-Jul-09
8435	http://www.stonetemple.com/articles/video-	http://www.	4-Jul-09
8436	http://www.stonetemple.com/articles/video-	http://www.	3-Jul-09
8437	http://www.stonetemple.com/blog	http://www.	9-Jul-09
8438	http://www.stonetemple.com/blog	http://www.	21-Jul-09
8439	http://www.stonetemple.com/blog	http://www.	2-Jul-09
8440	http://www.stonetemple.com/blog	http://www.	1-Jul-09
8441	http://www.stonetemple.com/blog	http://www.	14-Jul-09
8442	http://www.stonetemple.com/blog	http://www.	6-Jul-09
8443	http://www.stonetemple.com/blog	http://www.	10-Jul-09
8444	http://www.stonetemple.com/blog	http://www.	23-Jul-09
8445	http://www.stonetemple.com/blog	http://www.	22-Jul-09
8446	http://www.stonetemple.com/blog	http://www.	25-Jul-09
8447	http://www.stonetemple.com/blog	http://www.	28-Jul-09
8448	http://www.stonetemple.com/blog	http://www.	24-Jul-09
8449	http://www.stonetemple.com/blog	http://www.	20-Jul-09
8450	http://www.stonetemple.com/blog	http://www.	17-Jul-09
8451	http://www.stonetemple.com/blog	http://www.	26-Jul-09
8452	http://www.stonetemple.com/blog	http://www.	25-Jul-09
8453	http://www.stonetemple.com/blog	http://www.	24-Jul-09

You can then find the end of the links that go to the blog and see how many total links you have. So, for example, if the last link is found on line 12,367 of the spreadsheet, the total number of links to the blog is $12,367 - 8,437 = 3,930$. That can get you a raw count of inbound links to the blog. To go further, you can start analyzing what parts of the blog have the most links as well.

If you want additional data on the links to your blog, you can sign up to use Enquisite Optimizer, formerly known as Enquisite Optimizer, which will additionally allow you to track the traffic into your site from your links. You can organize it to track on a per-URL basis, on a per-domain basis, or even on a custom group of URLs or domains as a single item. Better still, you can also track the sales that result from such links.

Search Engine Robot Traffic Analysis

Understanding how robots are spidering your site is another thing that the expert SEO practitioner should know how to do. For one thing, spidering frequency is a clue as to which pages on your site have the highest PageRank and trust, because Google crawls the Web in reverse PageRank order. It can also help you detect spidering problems on your site.

You can use the tools we discuss in this section to help you find potential spidering issues, and analyze how important the search engines consider your content to be. You should be looking for clues of SEO problems, such as *robots.txt* blocking the crawlers, architectural problems, or even signs of a penalty (as might be signified by a big drop in crawling frequency).

However, it is important to know that this data will not tell you everything. For example, the fact that a web page was crawled does not mean it will be placed in the index. The page will still need to pass some additional tests to achieve such placement (such as the presence of unique content and enough links to justify its inclusion).

For pages that are indexed, you can look at how often the spiders visit/crawl your pages versus how often the engine actually shows a new version of your page in the index. To do this, look at the last cached date the search engine reported, and compare it with your crawling data.

Tools that perform log analysis include [Webtrends](#), [Unica Affinium NetInsight](#), and [Lyris HQ Web Analytics](#). These are well-known web analytics packages that offer the option of analyzing your logfiles. [Figure 9-41](#) shows a snapshot of a robot report from NetInsight.

N O T E

Google acquired Urchin in March 2005 (<http://www.google.com/intl/en/press/pressrel/urchin.html>). Urchin's JavaScript-based solution became Google Analytics.

However, the logfile analysis software version of Urchin continues to be available as well (<http://www.google.com/urchin/index.html>).

Other logfile analysis programs include:

- [Analog](#)
- [AWStats](#)
- [Sawmill](#)
- [Visitors](#)
- [Webalizer](#)
- [W3Perl](#)
- [Piwik](#)



FIGURE 9-41. NetInsight robot report

Google Webmaster Tools

You can also get detailed information about spidering activity on your own website using Google Webmaster Tools. Figure 9-42 shows a snapshot for one site.

This provides a great visual snapshot. One question that emerges from this data that this publisher may want to consider is why the time per page jumped up from 300 milliseconds to 650 milliseconds or so in mid-December. This may be the result of a change in the site architecture or a change in the hosting arrangements for the website. When you see these types of changes, it can be a flag that something happened with your website, and you should investigate it.

Web Traffic Comparison

When you set your goals for growing your online business, you may, for example, target a sales increase of 50% from search referral traffic. However, you may also want to take into account where your competitor is in setting those goals. If your competitor has twice your traffic, you may not be happy with a goal to increase your traffic by 50%.

In addition, when you measure the results of a completed SEO campaign, it is useful to see how much your competitor grew during the same time frame, because you can use this to set

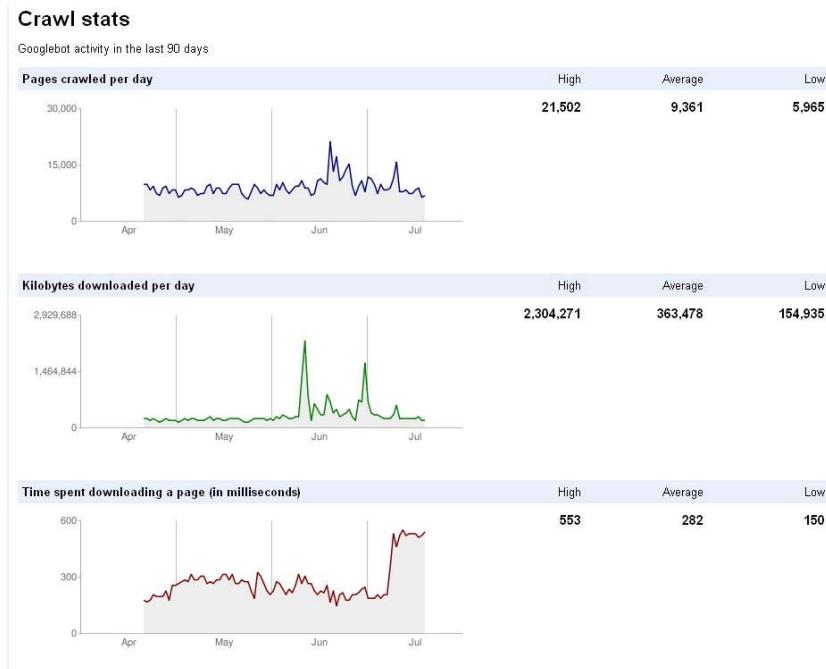


FIGURE 9-42. Google Webmaster Tools spider activity report

your goals for your next campaign. For example, if you grew your traffic by 50% and your competitor grew by 100%, you may want to adjust accordingly.

Several tools out there offer ways to compare the traffic of different websites. However, each tool is imprecise and has a significant bias.

For example, [Alexa](#) relies on data obtained from users who have the Alexa toolbar on their systems. This is a limited number of users, and of course, only certain types of people are willing to have the toolbar on their system. This introduces a bias to the data, and it means it is not completely reliable. In addition, you can get software to artificially inflate your Alexa rankings by spoofing the Alexa toolbar. This is not something you should do, but you should understand the issues with these types of measurement systems.

Similarly, companies such as [Compete](#) and [Quantcast](#) offer traffic measurement services. These are also subject to issues with bias, and limited sample sites.

However, these tools shine when comparing the traffic of one site to another. This is great data because if you are comparing two sites in the same market space (perhaps you and your competitor), the bias should affect both sites equally.

So, perhaps the tool will tell you that one site has twice the traffic as another one does. This type of relative measurement data is pretty accurate, and is very valuable information to get.

Google Trends for Websites

One tool that should not have the same level of bias is Google Trends for Websites.

[Figure 9-43](#) provides a snapshot of Google Trends for Websites showing the traffic comparison for [SearchEngineWatch.com](#) and [SearchEngineLand.com](#).

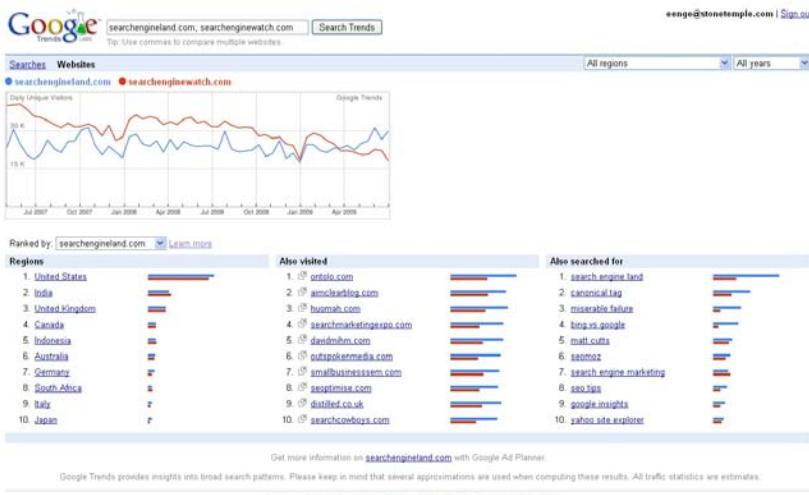


FIGURE 9-43. Google Trends for Websites

In addition to the comparative graph, you can see some details on the regions where the traffic came from, other sites visited by the same users, and related terms that people searched for. Note that the data Google provides is two to three months old.

Alexa

Alexa provides you with a quick and easy look at how the traffic for two sites compares, as shown in [Figure 9-44](#).

Underneath the chart you will notice a lot of additional data on [SearchEngineWatch.com](#) (the site listed first in the search box), such as the most popular content and where the readers are coming from.

Compete

Compete offers both a free and a paid service. As with Alexa, you can get basic traffic data on a site or compare traffic between sites (see [Figure 9-45](#)).



FIGURE 9-44. Alexa traffic comparison

Compete also offers additional data for a fee. This includes data on the major keywords driving traffic to a site, as shown in Figure 9-46.

Compete leverages multiple sources, including ISP, panel, and toolbar data, and has a total panel of about 2 million users. Compete's diverse data sources help in reducing bias in the data.

Quantcast

Quantcast is a competing service to the others we've listed. With Quantcast, you can get detailed data about individual sites, such as the data shown in Figure 9-47.



FIGURE 9-45. Compete site traffic comparison

The demographic data is interesting, as well as the other sites that users visit. Quantcast collects “pixel data” from scripts running on websites that are part of the program. Quantcast claims that these sites represent about 50 million visitors per month. Quantcast supplements this data with data from ISPs.

Quantcast also offers publishers an opportunity to participate in its Quantified program (<http://www.quantcast.com/docs/display/publisher/Publisher+Program+Overview>). This program enables publishers to make public much more accurate data through Quantcast (for publishers that want to leverage this data in setting their advertising pricing) and requires a tracking tag to be placed on their site.

Temporal Link Growth Measurements

Since Google released its patent application on historical information in March 2005 (read more at <http://www.seomoz.org/article/google-historical-data-patent>), search marketers have recognized that trends in temporal link and content analysis do have a real impact on rankings.

The engines are trying to measure patterns—they’re looking for indications of increasing or decreasing relevance and authority that temporal trends provide. They want to identify several specific items:

Content growth patterns

How often does a particular site tend to add new pages?

Content update patterns

How often are documents edited and updated?

Link growth patterns

How often do new links appear pointing to the site?

Link stagnation patterns

Does the number of links to the site stagnate or decrease?

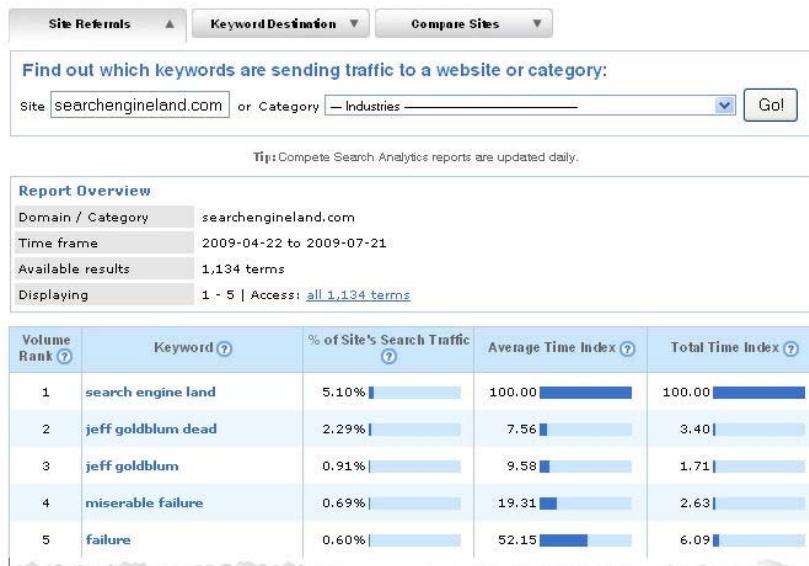


FIGURE 9-46. Compete keyword data

The engines aren’t interested only in how many links pointed to the site today versus yesterday (or how many pages were added); they are also fundamentally interested in tracking patterns over time. Figures 9-48 and 9-49 depict some example graphs showing the rate of new external links (and in the last two instances, pages) created over time, with some speculation as to what the trends might indicate.

These assumptions do not necessarily hold true for every site or instance, but the graphs make it easy to see how the engines can use temporal link and content growth information to make

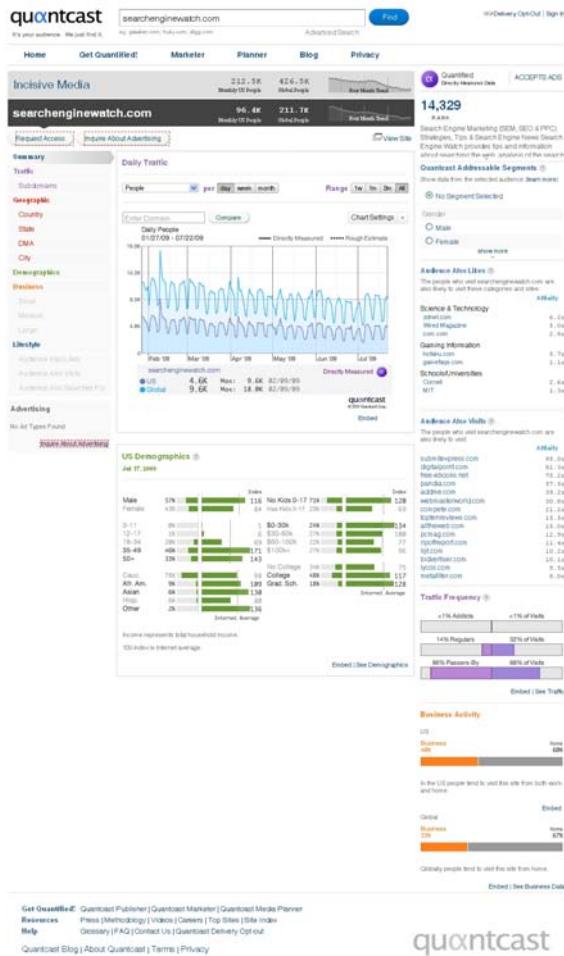


FIGURE 9-47. Quantcast site data

guesses about the relevance or worthiness of a particular site. Figure 9-50 shows some guesstimates of a few real sites and how these trends have affected them.

As you can see in Figure 9-50, Wikipedia has had tremendous growth in both pages and links from 2002 through 2006. This success manifests itself in the search engines, which reward Wikipedia's massive link authority with high rankings for much of its content.

Meanwhile, as shown in Figure 9-51, DMOZ has experienced a relative decline in popularity. Although DMOZ was once a default reference link for many sites, its relative influence has waned. It is likely that its traffic has declined in a similar fashion.

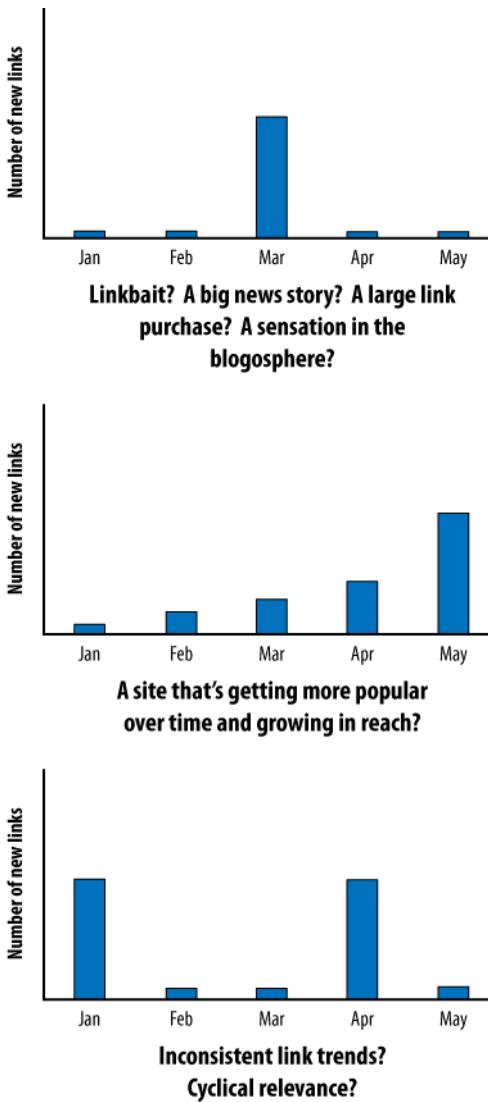


FIGURE 9-48. Interpreting new external link data

Many forms of spam and manipulative link building are likely to stand out like a sore thumb when put under the temporal microscope. When a large gain in links relative to a site's sphere, influence, and historical link growth appears, the engines can take a closer look at the source of the links or even trigger a manual review. Common sense would dictate that a small-time local real estate site doesn't usually attract a few thousand new links in a week unless it has done something newsworthy or linkworthy.

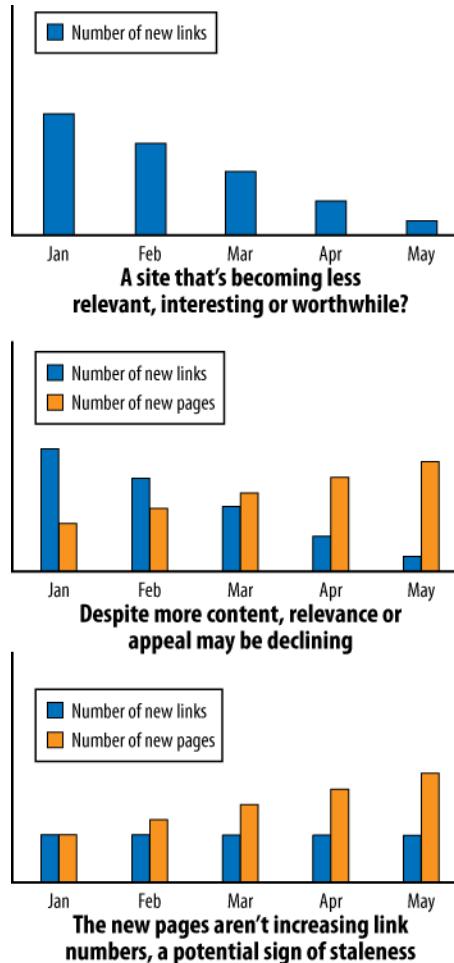


FIGURE 9-49. More link data speculation

There are few limits to what the engines can do with data such as this, and there is no reason they shouldn't be analyzing it (since it is easily available). Do consider how the link and content growth patterns for your sites may affect the engines' perspectives on your rankings and trustworthiness.

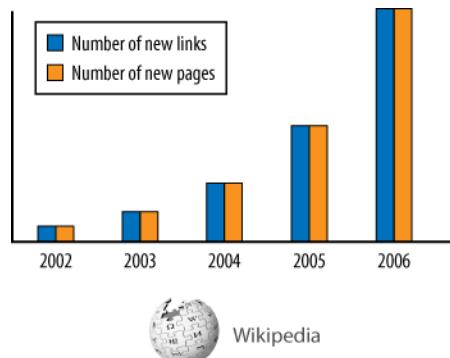


FIGURE 9-50. Wikipedia link data guesstimates

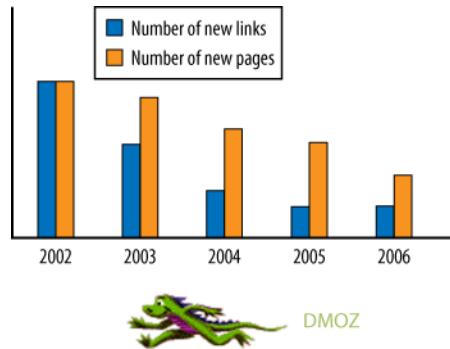


FIGURE 9-51. DMOZ link data guesstimates

Key Performance Indicators for Long Tail SEO

As we have discussed throughout the book, the long tail is an important part of SEO. Metrics are available for diagnosing the health of your long tail search traffic. Here are some that were developed by Brian Klais of Netconcepts.

Brand-to-Non-Brand Ratio

This is the percentage of your natural search traffic that comes from brand keywords versus non-brand keywords.

If the ratio is high and most of your traffic is coming from searches for your brand, this signals that your SEO is fundamentally broken. The lower the ratio, the more of the long tail of natural

search you likely are capturing. This metric is an excellent gauge of the success of your optimization initiatives.

Unique Crawled URLs

This is the number of unique (nonduplicate) web pages crawled by search engine spiders such as Googlebot. Your website is like your virtual sales force, bringing in prospects from the search engines. Think of each unique page as one of your virtual salespeople. The more unique pages you have, the more opportunities you have to sell through the search engines.

Search Visitors per Contributing Page

This is the percentage of unique pages that yield search-delivered traffic in a given month.

This ratio essentially is a key driver of the length of your long tail of natural search. The more pages yielding traffic from search engines, the healthier your SEO program is. If only a small portion of your website is delivering searchers to your door, most of your pages—your virtual salespeople—are warming the bench instead of working hard for you. You can think of these non-performing pages as “freeloaders.”

Keywords per Page

This is the average number of keywords each page (minus the freeloaders) yields in a given month. Put another way, it is the ratio of keywords to pages yielding search traffic.

The higher your keyword yield, the more of the long tail of natural search your site will capture. In other words, the more keywords each yielding page attracts or targets, the longer your tail. So, an average of eight search terms per page indicates pages with much broader appeal to the engines than, say, three search terms per page. The average online retailer in a Netconcepts study on the long tail of natural search (<http://www.netconcepts.com/learn/ChasingTheLongTail.pdf>) had 2.4 keywords per page.

Search Visitors per Keyword

This is the ratio of search-engine-delivered visitors to search terms. This metric indicates how much traffic each keyword drives and is a function of your rankings in the SERPs. Put another way, this metric determines the height or thickness of your long tail.

The average merchant in the aforementioned Netconcepts study obtained 1.9 visitors per keyword.

Index-to-Crawl Ratio

This is the ratio of pages indexed to unique crawled pages. If a page gets crawled by Googlebot, that doesn't guarantee it will show up in Google's index. A low ratio can mean your site doesn't carry much weight in Google's eyes.

Search Visitors per Crawled Page

Calculated for each search engine separately, this is how much traffic the engine delivers for every page it crawls.

Each search engine has a different audience size. This metric helps you fairly compare the referral traffic you get from each. The Netconcepts study found that Bing and Yahoo! tended to crawl significantly more pages, but the yield per crawled page from Google was typically significantly higher.

As you optimize your site through multiple iterations, watch the aforementioned KPIs to ensure that you're heading in the right direction. Those who are not privy to these metrics will have a much harder time capturing the long tail of SEO.

Conclusion

One of the biggest challenges for an SEO professional is the relatively long delay between implementation and results. Paid search campaigns can begin to show results in 24 hours, yet for SEO the process can take months. This makes effective measurement of results and progress critical. Putting the right analytics tools in place can provide the accountability required by budget managers. For the SEO practitioner, it can also provide valuable insight that can be used to tune and tweak campaigns for the best possible results.

CHAPTER TEN

Domain Changes, Post-SEO Redesigns, and Troubleshooting

WHENEVER YOU MAKE STRUCTURAL CHANGES TO YOUR WEBSITE, there is a risk that you will confuse the search engine and harm your search rankings. This can include changing your domain name, changing your content management system (CMS), redesigning your site—anything that fundamentally alters the landscape of your web presence.

In this chapter, we will review these types of scenarios from the top down. Be ready to refer back to “[Content Delivery and Search Spider Control](#)” on page 238 in [Chapter 6](#) to review the technical specifics of options for moving content from one location to another.

The Basics of Moving Content

Moving content refers to the situation when content used to live at one URL (e.g., *http://www.yourdomain.com/pageA.html*) and has now moved to another URL (e.g., *http://www.yourdomain.com/products/pageA.html*). One of your goals when you move content is to make sure users (and search engines) that attempt to visit the old URL (*/pageA.html*) are presented with the content from the new location (*/products/pageA.html*).

It turns out that there is an additional critical goal: namely, when you move content from one URL to another, the links to that content at the old URL will stop providing value to your rankings in the search engines for that content, unless you perform a 301 redirect.

In “[Duplicate Content Issues](#)” on page 226 in [Chapter 6](#), we covered the technical specifics of how to do this in detail, including the golden rule of moving content: the search engine needs to see a 301 HTTP status code whenever you redirect it (and users) to a new location.

The 301 HTTP status code causes the search engine to pass most of the value of any links the original page has over to the new page, and should result in the rapid de-indexation of the old URL. Since link juice is a previous asset, you want to make sure you use a 301 redirect every time!

Large-Scale Content Moves

The 301 process can become difficult when changes result in movement of large quantities of content. For example, when you change your domain name, every single piece of content on your site will move to a new URL, even if the site architecture is identical (*http://www.olddomain.com/...* moves to *http://www.newdomain.com/...*).

This is challenging because you might have to set up individual 301 redirects for every single page on the site, as in this example:

```
http://www.olddomain.com/page1.html 301 redirect to http://www.newdomain.com/page1.html
http://www.olddomain.com/page2.html 301 redirect to http://www.newdomain.com/page2.html
http://www.olddomain.com/page3.html 301 redirect to http://www.newdomain.com/page3.html
...
...
http://www.olddomain.com/page1000.html 301 redirect to http://www.newdomain.com/
page1000.html
```

In this simple example, we would have to set it up so that all 1,000 pages on the old domain redirect to the same content on the new domain. In some systems, these need to be set up one at a time, so it could be quite painful. Imagine a site with 1 million pages!

However, publishers who use an [Apache](#) web server (Unix and Linux servers) can take advantage of the power of Apache’s [mod_rewrite module](#), which can perform the redirect of every URL on the old domain to the same URL on the new domain in two lines of code:

```
RewriteCond %{HTTP_HOST} ^olddomain\.com [NC]
RewriteRule ^/(.*) http://www.newdomain.com/$1 [R=301,L]
```

The preceding code presumes that you prefer the “www” version as the canonical URL. You can also use two similar lines of code to specify the “non-www” version as the canonical URL (see <http://hamletbatista.com/2007/07/19/canonicalization-the-gospel-of-http-301/> for examples without “www” and other alternative approaches).

When you are moving a site with lots of pages from one domain to another, that’s a beautiful thing! You can also use this sort of scalable approach for other large-scale content moves. The details of the instructions depend on the exact nature of the content move you are making.

The other highly popular web server is [IIS](#) from Microsoft. In many installations of IIS, you will find yourself in a situation where you have to implement a separate instruction for each page, one at a time.

However, rewrites can be done on IIS too, with the help of an ISAPI plug-in such as [ISAPI_Rewrite](#). When this is installed you can perform large scalable rewrites in a language similar to that used by Apache's mod_rewrite. You can learn more about mod_rewrite, ISAPI_Rewrite, and regular expressions in "[Redirects](#)" on page 190 in [Chapter 6](#).

Mapping Content Moves

The first stage of dealing with a site redesign is to figure out which content will be moved where and which content will be removed altogether. You will need this to tell you which URLs you will need to redirect and to which new locations.

To do this you must start by getting a complete map of your URLs. For many websites this is not as simple as it sounds. Fortunately, tools are available to make the job easier. Here are some ways to tackle this problem:

- Extract a list of URLs from your web server's logfiles.
- Pull the list from your XML Sitemap file, provided you believe it is reasonably complete.
- Use a free crawling tool such as Xenu or GSiteCrawler.
- Use Google Webmaster Tools to pull a list of the external links to your site, and make sure all pages that have received links on your site are included.

These tools should help you assemble a decent list of all your URLs. One way to do this is to lay it out in a spreadsheet, which might end up looking like [Table 10-1](#).

TABLE 10-1. Planning your content moves in advance

http://www.olddomain.com/page1.html	http://www.newdomain.com/page1.html
http://www.olddomain.com/page2.html	http://www.newdomain.com/page2.html
http://www.olddomain.com/page3.html	http://www.newdomain.com/page3.html
http://www.olddomain.com/page4.html	http://www.newdomain.com/page4.html
http://www.olddomain.com/page5.html	http://www.newdomain.com/page5.html
http://www.olddomain.com/page6.html	http://www.newdomain.com/page6.html
http://www.olddomain.com/page7.html	http://www.newdomain.com/page7.html
http://www.olddomain.com/page8.html	http://www.newdomain.com/page8.html
http://www.olddomain.com/page9.html	http://www.newdomain.com/page9.html
http://www.olddomain.com/page10.html	http://www.newdomain.com/page10.html

If you are redirecting a massive number of URLs, you should look for ways to simplify this process, such as writing rules that communicate what you need to know. We could abbreviate the list in [Table 10-1](#) to the short list in [Table 10-2](#).

TABLE 10-2. Simplifying content move planning with wildcards

http://www.olddomain.com/page*.html	http://www.newdomain.com/page*.html
---------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------

Then you can save the individual lines for the more complicated moves, so your resulting spreadsheet would look like [Table 10-3](#).

TABLE 10-3. Mapping all your content moves completely

Individual page moves	
http://www.olddomain.com/about-us.html	http://www.newdomain.com/about-us.html
http://www.olddomain.com/contact-us.html	http://www.newdomain.com/contact-us.html
http://www.olddomain.com/press-relations.html	http://www.newdomain.com/press.html
Large-scale page moves	
http://www.olddomain.com/content/*.html	http://www.newdomain.com/content/*.html
http://www.olddomain.com/page*.html	http://www.newdomain.com/page*.html

The purpose of this is not to write the code for the developers, but to efficiently give them a map for how the content movement should take place. Note that the spreadsheet should contain a map of all changed URLs, which may include downloadable content such as PDF files, PowerPoint presentations, Flash files, multimedia, or any other such content that is being moved.

You should also note which content will no longer exist. You can do this as additional entries in the left column, with the entry in the right column indicating where users looking for that old content should be remapped. Now your spreadsheet might look like [Table 10-4](#).

TABLE 10-4. Identifying pages that have been removed

Individual page moves	
http://www.olddomain.com/about-us.html	http://www.newdomain.com/about-us.html
http://www.olddomain.com/contact-us.html	http://www.newdomain.com/contact-us.html
http://www.olddomain.com/press-relations.html	http://www.newdomain.com/press.html
Large-scale page moves	
http://www.olddomain.com/content/*.html	http://www.newdomain.com/content/*.html
http://www.olddomain.com/page*.html	http://www.newdomain.com/page*.html

Eliminated pages	
http://www.olddomain.com/widgets/azure	http://www.newdomain.com/widgets/blue
http://www.olddomain.com/widgets/teal	http://www.newdomain.com/widgets/green
http://www.olddomain.com/widgets/puce	http://www.newdomain.com/widgets/

The new entries show what should happen to eliminated pages. The first two eliminated pages may represent products that you no longer carry, so you redirect them to the closest existing product you have. The third eliminated page represents one where there is no direct good fit, so we chose that we want to redirect that one to the parent page for that topic area.

Ultimately, the reason for this detailed mapping is that we want to preserve as much link juice from the old URLs as possible while providing the best user experience for people who arrive at the old URLs.

Expectations for Content Moves

The big downside to all this is that the search engines won't necessarily adapt to these moves instantaneously. Many sites temporarily lose rankings after making a large-scale content move, then recover after a period of time. So naturally, the question is, how long will it take to get your rankings and traffic back?

The reality is that a number of factors are involved, depending on your particular situation. Some examples of these factors might include:

Size and complexity of your site

Bigger, more complex sites may take longer to process.

Complexity of the move

If the site has been fundamentally restructured, it is likely to take more time for the search engines to adapt to the new structure.

Perceived authority of the site

Sites that have a higher (search engine) perceived authority may be processed faster.

Related to this is the rate at which the site is typically crawled.

The addition of new links to the new pages

Obtaining new links to the new URLs, or changing old links that used to point to the old URLs so that they point to the new URLs, can help speed up the process.

If you are moving to an entirely new domain, you can aid the process in Google by using the Change of Address tool inside Google Webmaster Tools. Before using this tool make sure that both your old domain and your new domain are verified in Webmaster Tools. Then, on the Webmaster Tools home page, click on the old domain and under Site Configuration click "Change of address." Then select the new site.

When all is said and done, a reasonable estimate would suggest that a significant traffic dip from the search engines should rarely last longer than 60 to 90 days, and many are picked up in a dramatically shorter time span.

Maintaining Search Engine Visibility During and After a Site Redesign

Companies may decide to launch a site redesign as part of a rebranding of their business, a shift in their product lines, a marketing makeover, or for any number of reasons. During a site redesign, any number of things may change on the site. For example:

- Content may move to new URLs.
- Content might be eliminated.
- Content could be moved behind a login.
- New sections may be added.
- New functionality may be added.
- Navigation/internal linking structure may be changed significantly.

Of course, this may involve moving everything to a new domain as well, but we will cover that in “[Maintaining Search Engine Visibility During and After Domain Name Changes](#)” on page 445. Here are some best practices for handling a site redesign:

- Create 301 redirects for all URLs from the original version of the site pointing to the proper URLs on the site. This should cover scenarios such as any remapping of locations of content and any content that has been eliminated. Use a spreadsheet similar to the one we outlined at the beginning of this chapter to map out the moves to make sure you get them all.
- Review your analytics for the top 100 or so domains sending traffic to the moved and/or eliminated pages, and contact as many of these webmasters as possible about changing their links. This can help the search engines understand the new layout of your site more rapidly and also provides a better branding and user experience.
- Review a backlink report (using your favorite backlinking tool) for your site and repeat the process in the preceding bulleted item with the top 200 to 300 or so results returned. Note that reports from Yahoo! Site Explorer include NoFollowed links, so there is a certain degree of error in this approach with Site Explorer. Consider using more advanced tools, such as [Linkscape](#) or [Majestic-SEO](#), which allow you to filter your links to more easily identify the most important ones.
- Make sure you update your [Sitemap](#) and submit it to [Google’s Webmaster Central](#).
- Monitor your rankings for the content, comparing old to new over time—if the rankings fall, post in the [Google Groups Webmaster Central Forum](#) information regarding what you did, what happened, and any information that might help someone help you.

- Monitor your Webmaster Central account for 404 errors and to see how well Google is doing with your 301s. When you see some 404s pop up, make sure you have a properly implemented 301 redirect in place. If not, fix it. Don't limit this checking just to 404 errors. Also be on the lookout for HTTP status codes such as 500, 302, and others.

Maintaining Search Engine Visibility During and After Domain Name Changes

Sometimes when you do a site redesign or a company rebranding, you also change your domain name. But sometimes publishers are simply changing the domain name, and everything else (except for perhaps branding changes) stays the same.

Unique Challenges of Domain Name Changes

One of the more challenging aspects of a domain name switchover is the potential loss of trust the search engines attached to the old domain. The trust does not always shift that easily to the new domain when you make the move. Another issue is that the keywords present in the old domain name but not in the new domain name may negatively impact for search terms that include the keyword. For example, when Matt Mullenweg, founder of WordPress, switched domains on his personal site from <http://photomatt.net> to <http://ma.tt>, his position in Google for *Matt* dropped out of the #1 spot.

One danger is that the new domain may go through a scenario where the newness of the domain acts as a damper on the site's rankings. In other words, the site relevance and inbound link profile may suggest a high ranking for some search queries, but because the site is not trusted yet, the rankings are suppressed, and traffic is much lower than it would otherwise be.

This is one reason for the recommendation we will make in the upcoming best practices list to get the most important links to your old domain switched over to your new domain. One other tactic you can try is to make the move to a different domain that has a history associated with it as well. Just make sure the history is a positive one! You don't want to move to an old domain that had black marks against it. Of course, the tactic of moving to an old domain is viable only if it is compatible with the reasons you are making the domain move in the first place.

Unfortunately, lost traffic is common when you make such a move. If you do all the right things, you can and should recover, and hopefully quickly. You should, however, be prepared for the potential traffic impact of the switchover.

Premove Preparations

If you are using a new domain, buy it as early as you can, get some initial content on it, and acquire some links. The purpose of this exercise is to get the domain indexed and recognized by the engines ahead of time (to help avoid or shorten sandboxing).

Then, register the new domain with Google Webmaster Central, Bing Webmaster Tools, and Yahoo! Site Explorer. This is just another part of making sure Google knows about your new domain as early as possible and in as many ways as possible.

Once this is done, here are the best practices for handling a domain name change:

- Create 301 redirects for all URLs from the old site pointing to the proper URLs on the new site. Hopefully you will be able to use mod_rewrite or ISAPI_Rewrite to handle the bulk of the work. Use individual rewrite rules to cover any exceptions. Have this in place at launch.
- Review your analytics for the top 100 or so domains sending traffic to the old pages, and contact as many of these webmasters as possible about changing their links.
- Review a Yahoo! Site Explorer report for your site and repeat the process in the preceding bulleted item with the top 200 to 300 results returned.
- Make sure that both the old site and the new site have been verified and have Sitemaps submitted at Google Webmaster Central, Bing Webmaster Tools, and Yahoo! Site Explorer.
- Launch with a media and online marketing blitz—your goals are to get as many new inbound links pointing to the site as quickly as possible, and to attract a high number of branded searches for the redesigned site.
- Monitor your rankings for the content, comparing old to new over time. If the rankings fall, post in your thread at Google Groups with an update and specifics.
- Monitor your Webmaster Central account for 404 errors and to see how well Google is doing with your 301s. When you see some 404s pop up, make sure you have a properly implemented 301 redirect in place. If not, fix it.
- Monitor the spidering activity on the new domain. This can provide a crude measurement of search engine trust. Search engines spend more time crawling sites they trust. When the crawl level at the new site starts to get close to where it was with the old site, you are probably most of the way there.
- Watch your search traffic referrals as well. This should provide you some guidance as to how far along in the process you have come.
- You can also check your server logs for 404 and 500 errors. This will sometimes flag problems that your other checks have not revealed.

An additional idea comes from Matt Cutts, who suggested the following at PubCon 2009 (<http://blog.milestoneinternet.com/web-development/faq-on-duplicate-content-and-moving-your-site-by-matt-cutts-at-pubcon-2009/>):

So here's the extra step. Don't just move the entire domain from the old domain to the new domain. Start out and then move a sub-directory or a sub-domain. Move that first; if you've got a forum, move one part of your forum. Move that over to the new domain, and make sure that the rankings for that one part of your site don't crash. Sometimes it takes a week or so for them to sort of equalize out, because we have to crawl that page to see that it's moved. So if you move

a part of your site first, and it goes fine, then you know that you're pretty safe. So instead of doing one huge move, if you can break it down into smaller chunks and start out by moving a small part of your site first, you'll know that you'll be gold.

The value of this approach is that it reduces the risk associated with the move into more manageable chunks. Even if you use this approach, however, you should still follow the process we outlined in this section to implement the move of each site section, and check on its progress.

Changing Servers

Sometimes you may decide you want to move servers without changing your domain name or any of your URLs. A common cause of this is that the growth of your traffic requires you to move up in terms of hosting environment. If you are using third-party hosting, perhaps you are changing your hosting company. If you have your own data center, you may need to move or expand your facilities, resulting in a change in the IP addresses of your servers.

This is normally a straightforward process: simply go to the registrar where you registered the domain name and update the DNS records to point to the new server location. You can also temporarily decrease the site's DNS Time to Live (TTL) to five minutes (or something similar) to make the move take place faster. For the most part, you should be done, though you should follow the monitoring recommendations we will outline shortly.

Even if you follow this process, certain types of problems can arise anyway. Here are the most common ones:

- You may have content that can't function on the new platform. A simple example of this might be that you use Perl in implementing your site, and Perl is not installed on the new server. This can happen for other reasons, and these can result in pages that return 404 or 500 errors, instead of the content you intended.
- Unfortunately, publishers commonly forget to move key content or files over, such as *robots.txt*, analytics files, *sitemaps.xml*, the *.htaccess* file, and so forth. The first tip, of course, is not to forget to move these files, but people are human and they do make mistakes.
- Server configuration differences can also lead to mishandling of certain types of requests. For example, even if both your old server and your new server are running IIS, it is still possible that the new server is configured in such a way that it will turn any 301 redirects you have in place to 302 redirects, which would be unfortunate indeed!

The best advice for dealing with these concerns is to make a list of special files and configuration requirements and verify that everything is in place prior to flipping the switch on the server move.

In addition, you should conduct testing of the new site in its new location before flipping the switch. You will need to access the content on the new site using its physical IP address. So,

the page at *http://www.yourdomain.com/pageA.html* will be found at an address similar to *http://206.130.117.215/pageA.html*. To do this, you should add that IP address to your test machine’s *hosts* file (this assumes you are running Windows) with a corresponding hostname of *http://www.yourdomain.com*, which will allow you to surf the site at the new IP address seamlessly. This advance testing should allow you to check for any unexpected errors. Note that the location of the *hosts* file varies across different versions of Windows, so you may need to search online to get information on where to find it on your machine.

Monitoring After Your Server Move

As with our other scenarios, post-launch monitoring is important. Here are the basic monitoring steps you should take:

- Monitor your Webmaster Central account for 404 errors and to see how well Google is doing with your 301s. When you see some 404s pop up, make sure you have a properly implemented 301 redirect in place. If not, fix it.
- Monitor the spidering activity on the new domain to make sure no unexpected drops occur.
- Watch your search traffic referrals for unexpected changes.
- You can also check your server logs for 404 and 500 errors. This will sometimes expose problems that your other checks have not revealed.

Other Scenarios Similar to Server Moves

Sometimes organizations transition from a single-server environment to a multiserver environment because of increasing levels of traffic (a good problem to have!). How this affects them depends on how they make the transition. The wrong way to do it is to have one server at *http://www.yourdomain.com*, another at *http://www2.yourdomain.com*, yet another at *http://www3.yourdomain.com*, and so forth. This is because all these scenarios are creating duplicate content. Unfortunately, this type of mistake is common. If you search on *inurl:www2* in Google, it returns 346 million results! (See [Figure 10-1](#).)

Fortunately, when you are knowledgeable about that issue, it is usually not too difficult to implement a transparent solution in which the URL always starts with “www” regardless of the server providing it.

Hidden Content

Sometimes publishers produce great content and then for one reason or another fail to expose that content to search engines. In [“Content Delivery and Search Spider Control” on page 238](#) in [Chapter 6](#), we discussed ways that you can hide content from the search engines when you want to. However, at times this is done unintentionally. Valuable content can be inadvertently

Google search results for "inurl:www2". The results are as follows:

- Shuttle Radar Topography Mission**
This is the SRTM home page. The Shuttle Radar Topography Mission (SRTM) obtained elevation data on a near-global scale to generate the most complete ...
www2.jpl.nasa.gov/srtm/ - Cached - Similar -
- Oregon Scientific IUS**
Oregon Scientific is a designer of award-winning, upscale personal electronic products in the consumer electronics and toy industries including weather ...
www2.oregonscientific.com/ - Similar -
- Goldman Sachs**
A full-service global investment banking and securities firm. [Show stock quote for GS](#)
www2.goldmansachs.com/ - Cached - Similar -
- The Education Trust - Closing the Achievement Gap**
The Education Trust works for the high academic achievement of all students at all levels, kindergarten through college, and forever closing the achievement ...
www2.edtrust.org/ - Cached - Similar -

FIGURE 10-1. Search results for “inurl:www2”

hidden from the search engines, and occasionally, the engines can find hidden content and construe it as spam, whether that was your intent or not.

Identifying Content That Engines Don’t See

How do you determine when this is happening? Sometimes the situation is readily apparent; for example, if you have a site that receives high traffic volume and then your developer accidentally `NoIndexes` every page on the site; you will begin to see a catastrophic drop in traffic. Most likely it will set off a panic investigation, leading to the `NoIndex` issue as the culprit.

Does this really happen? Unfortunately, it does. Here is an example scenario. You work on site updates on a staging server. Because you don’t want the search engines to discover this duplicate version of your site, you keep the pages on the staging server `NoIndexed`. Then, when someone moves the site from the staging server to the live server, he forgets to remove the `NoIndex` tags. It is just normal human error in action.

This type of problem can also emerge in another scenario. Some webmasters implement a `robots.txt` file that prohibits the crawling of their staging server website. If this file gets copied over when the site on the staging server is switched to the live server, the consequences will be just as bad as in the `NoIndex` scenario we just outlined.

The best way to prevent this type of scenario is to introduce a series of safety checks on the site that take place immediately after any update of the live server.

There are potential problems, however, that are much more difficult to detect. First, with a new site launch, you won’t have any preexisting traffic, so there will be no drop in traffic levels to flag that something is wrong. In another scenario, you may have an established site where you accidentally do something to hide only a portion of the site from the engines, so the issue is less obvious.

Regardless of your situation, web analytics can help you in the detection process. Use your analytics software to find pages on your site that get page views but no referring search traffic. By itself, this is not conclusive, but it provides a leading clue as to where to start. Note that the converse of this is interesting for another situation—if you see something that has search referrals and you don't want it to, you may want to hide that content.

Figure 10-2, from Netconcepts' GravityStream (<http://www.gravitystream.com>), shows pages not receiving search traffic during the specified time period, sorted by crawler activity. The report helps identify low-hanging fruit for the SEO practitioner to focus on. With zero searchers coming in, there's nowhere to go but up. Furthermore, more interest from Googlebot suggests higher importance/trust/authority, and thus greater potential to rank once the page has been optimized.

The screenshot shows the GravityStream interface with the following details:

- Header:** System Date: July 16th, 2009 | Current User: sspencer@netconcepts.com | Settings | Log Out
- Left Sidebar:**
 - DASHBOARD
 - REPORTS
 - ALL PAGES REPORT
 - PERFORMING PAGES
 - NON-PERFORMING PAGES**
 - SEARCH PHRASE REPORT
 - REFERRAL REPORT
 - CRAWLER SUMMARY
 - SEARCH ENGINE REPORT
 - LOCATION REPORT
 - RESPONSE CODES
 - DAILY BILLABLE TRAFFIC
 - OPTIMIZATION SUMMARY
 - All Optimizations
 - Awaiting Approval
- Main Content:**

Non-Performing Pages

DATE: June 2008 | PROGRAM: Select

Search Per Page: 20

June 01, 2009 - June 30, 2009 | Displaying 1 to 20 of 226,158 results

Page Path	Crawler	Optimized	View	Edit
/popular-keywords.shtml	54			
/popular-tractors.html	36			
/popular/schools.html	35			
/beauty-therapy-weight-reduction.shtml	35			
/popular/cafes.html	34			
/sitemap.html	29			
/schools-divers-commercial.shtml	24			
/automotive-repairs-car-truck-dealers.shtml	23			
/browse/companies/	23			
/popular/food.html	23			
/cafes-coffee-merchants.shtml	23			
/hairdressers-east-lazy.shtml	23			
/lawyers-ingewood-area.shtml	22			
/popular/category+e.html	22			
/popular/accommodation.html	22			
/popular/category+a.html	22			
/popular/category+n.html	21			
/florists-springfield-florists.shtml	21			
/popular/category+o.html	21			
/popular/category+r.html	21			
Totals:	534			
Grand Totals:	394,565			

Pagination: 1 2 3 4 5 6 7 8 9 next>> Jump 10>> last>>

© 2009 GravityStream 2.0. All Rights Reserved.

FIGURE 10-2. GravityStream report showing pages with no search traffic

Another data point you can examine is the number of pages the search engines report as indexed for your site. In a new site scenario, you can see whether the search engines appear to be picking up your content. For example, if you have a site with 1,000 pages with a good inbound link profile, and after three months only 10 pages are indexed, that could be a clue that there is a problem.

You do need to be careful not to overreact to the count of indexed pages, because the numbers that the search engines report will naturally fluctuate quite a bit. But if you are aware of the types of numbers typically reported for your site, and they drop to an unusually low level and stay there (or keep dropping), you probably have a problem.

Identifying the Cause of Non-Spidering

Once you realize you have a problem, you will need to move on to figuring out what is causing it. Here are some of the most common causes.

Blocked by robots.txt

Check your *robots.txt* file (<http://www.yourdomain.com/robots.txt>) to see whether you are preventing the crawlers from accessing parts of the site that you actually want them to see. This mistake is quite common.

Both Google Webmaster Tools and Bing Webmaster Tools provide simple ways to see whether they are aware of content that *robots.txt* is blocking them from crawling (see Figures 10-3 and 10-4).

FIGURE 10-3. Google Webmaster Tools: restricted by robots.txt

These reports are helpful when you have content on the site that has links to it (either internal or external), but the search engines are excluded from it in *robots.txt*, so they don't crawl it.

The solution is simple. Figure out what line in your *robots.txt* file is blocking the search engines and remove it, or adapt it so that the content is no longer blocked.

Blocked by the robots meta tag

The *robots* meta tag might look something like this:

```
<meta name="robots" content="noindex,nofollow">
```

The screenshot shows the 'Crawl Issues' tab selected in the navigation bar. Below it, a section titled 'Crawl issues' with a question mark icon is displayed. A note explains that the tool helps track down missing or broken links. A filter allows excluding results from specific domains. A dropdown menu under 'Select Issue Type' has 'File Not Found (404)' selected. Other options include 'Blocked By REP', 'Long Dynamic Url', 'Malware Infected', and 'Unsupported Content-Type'. There are also 'Show all selected' and 'Filter selected' radio buttons, and a 'Search' button.

Results for File Not Found (404)
Bing encountered a "404 File Not Found" HTTP status code when last attempting to download these URLs. This could either mean the page has been removed or somewhere on the web pointing to your site. We recommend you use this report to find a list of URLs to either repair the broken link, or create a 301 redirect. [Learn more](#)

FIGURE 10-4. *Bing Webmaster Tools: blocked by REP*

As we discussed in “[Content Delivery and Search Spider Control](#)” on page 238 in [Chapter 6](#), a setting of `NoIndex` will tell the search engine that it is not allowed to include the page in its index. Clearly, you should check to see whether you have made this error.

In addition, the `NoFollow` tag will tell search engines not to pass any link juice to the pages linked to on that page. If all the links on your site to a particular piece of content are `NoFollowed`, you are passing no link juice to the page. This tells the search engines that you don’t value them, and as a result they won’t bother indexing that content (unless someone else links to it without using `NoFollow`).

Solving this requires locating the places where these robots meta tags are on your site and removing them. Note that the default setting for the robots meta tag is “`index, follow`”, so there is no need to implement the tag if that is your desired setting. Just make sure you don’t have robots meta tags in place that change the default.

No direct links

You may find that a particular piece of content has no links, or that all of the links to that content are `NoFollowed`. You can also make links invisible to the search engines by showing them only to visitors that support JavaScript or cookies, or by encrypting the links to the content in some fashion. The solution here is to make sure you implement plain text (or image) links to the content. Better still is if you can get some third-party websites to link to the content as well.

Requires form submission

Requiring a login or some other type of form submission to see content is another common cause of nonspidering. Search engines will not attempt to fill out forms to see what is behind

them. The simplest solution can be to remove the requirement for the form if you want the search engines to index it.

However, some sites sell content on a subscription basis, and they will not want to offer their content for free. In October 2008, Google announced the First Click Free concept (<http://googlewebmastercentral.blogspot.com/2008/10/first-click-free-for-web-search.html>) to allow subscription-based sites to have their content crawled and indexed by Google, but still allow the publisher to require human visitors to pay to access the content. You can read more about First Click Free in “Content Delivery and Search Spider Control” on page 238 in Chapter 6.

Session IDs

Session IDs confuse search engine crawlers. Every time they come to your site, they see a different page. For example, they may see <http://www.yourdomain.com?SessID=2143789> one time and <http://www.yourdomain.com?SessID=2145394> the next time. Even though your intent is to track the session of a particular user, and you think of these as the same page, the search engine does not. You can read more about session IDs in “Controlling Content with Cookies and Session IDs” on page 234 in Chapter 6.

Not enough link juice to remain in main indexes

Sometimes the problem has nothing to do with the issues we just discussed. Search engines may see the page just fine, but there is not enough link juice going to it to merit inclusion in their main index. This is more common than people think, but the search engines do not attempt to index all the world’s web pages.

For example, content that Google perceives to be of low importance (i.e., not enough link juice) will be excluded from the main index, and will instead be placed in Google’s supplemental index or not indexed at all. This is because Google wants to emphasize the more important pages on the Web. Google takes this approach because it doesn’t want the performance with which it delivers search results to be slowed down by pages that most people probably don’t want to see.

The solution is to make the pages more important in the eyes of the search engines by adding internal links and campaigning to get external links to the content. Another related scenario is where the pages get screened out because the search engines perceive them to be duplicate content. The best solution for this version of the problem is to add unique content to those pages.

Hidden Content That May Be Construed As Spam

Hidden text is one of the challenges that webmasters and search engines face. Spammers continue to use hidden text to stuff keywords into their pages for purposes of artificially boosting their rankings. Search engines seek to figure out when spammers are doing this and then take appropriate action. There are many ways to create hidden text unintentionally, and

no one wants to be penalized for something they did not intend to do. To understand this better it is useful to examine Google's Webmaster Guidelines for hidden text (<http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=66353>) to see the bottom line:

If your site is perceived to contain hidden text and links that are deceptive in intent, your site may be removed from the Google index, and will not appear in search results pages.

In a [Google Groups thread](#), Googler Susan Moskwa had this to say:

Of course, as with many techniques, there are shades of gray between ‘this is clearly deceptive and wrong’ and ‘this is perfectly acceptable’. Matt (Cutts, Google’s Head of Web Spam) did say that hiding text moves you a step further towards the gray area. But if you’re running a perfectly legitimate site, you don’t need to worry about it. If, on the other hand, your site already exhibits a bunch of other semi-shady techniques, hidden text starts to look like one more item on that list. It is like how 1 grain of sand isn’t noticeable, but many grains together start to look like a beach.

Related to this is a posting by Matt Cutts on [Threadwatch](#):

If you’re straight-out using CSS to hide text, don’t be surprised if that is called spam. I’m not saying that mouseovers or DHTML text or have-a-logo-but-also-have-text is spam; I answered that last one at a conference when I said “imagine how it would look to a visitor, a competitor, or someone checking out a spam report. If you show your company’s name and it is Expo Markers instead of an Expo Markers logo, you should be fine. If the text you decide to show is ‘Expo Markers cheap online discount buy online Expo Markers sale ...’ then I would be more cautious, because that can look bad.”

Obviously, this is a fate you want to avoid. Note the use of the word *perceived* in the Google Webmaster Guidelines snippet. It doesn’t sound like a simple black-and-white problem, does it? In fact, it is not, as there are many ways to create hidden text.

A few ways to create hidden text

There are many techniques for creating hidden text. You can perform some of these without the use of CSS, and they are usually fairly easy to detect:

- Make your text and background colors identical (or virtually identical). This is the original method spammers used for creating hidden text. It is easy to detect.
- Set the font size for text to 0 or to a negative number. This is also easy to detect.
- Use a `noscript` tag. This is the tag that tells the browser what to put in place on the page when there is a block of JavaScript on the page but the browser has JavaScript support disabled. This is really only “pseudohidden text.” Although it is possible to make the text contained within the `noscript` tags different from what is in the JavaScript, about 3% of users will see it, and that’s more than enough to generate spam report complaints to the search engines. In other words, stuffing a lot of keywords within `noscript` tags comes with a fair amount of risk.

- Place your text way below the fold. This is also a pseudohidden text technique: that of providing content that is really not there for users. So, although it is visible, the text is clearly out of the action-oriented area of the page and resides well below the fold, and the user needs to scroll down to see it. The text could be directly related to the site's basic purpose, and the intent in this case would be that of keyword stuffing. It is hard to detect algorithmically, but under human review, it would likely be seen as a poor-quality signal.
- Use CSS-based methods for hiding text. CSS techniques for creating hidden text are more challenging because they are much harder for search engine crawlers to detect unless they crawl and interpret the CSS. Most crawlers don't do that currently. Many such techniques are used for completely legitimate purposes—for example, the `display:none` attribute is often used for creating drop-down menus.

Unintentionally creating hidden text

There are a few ways to create hidden text without intending to do so. One of the most common ways is via your CMS, which has some CSS-based methods built into it. For example, many CMSs use the `display:none` technique to implement drop-down menus, or other widgets that the user clicks on that then “expand” to display more text. Tab folders are a great example of this. Sometimes the `display:none` technique is used in user-generated content systems where the page normally shows the number of comments on a post, but chooses to suppress the text “0 Comments” in the event that no comments have been made yet.

Another common way that people create hidden text occurs when they start providing enhancements for the visually impaired. As with the example provided earlier of using hidden labels within a table, this comes about because you are trying to place text in a place that will make it look cluttered to a user with normal vision. The solution some people use to serve both audiences is to hide the text from the sighted users.

Many of these scenarios have no SEO value, even when manipulated by spammers. These types of techniques generally do not carry a risk of being penalized, because there is no reason to suspect negative intent.

Spam Filtering and Penalties

Over time, it has become a lot more difficult to “game” the search engines and a lot easier to fall victim to a search engine penalty or outright ban. It is hard to recover from these.

Consequences can include ranking penalties, removal of the site’s “voting” power (i.e., ability to pass PageRank), incomplete indexation (i.e., a partial site ban), or, worst of all, a total site ban.

Not even the largest corporations spending big dollars on Google AdWords are immune. For example, BMW had its entire BMW.de site banned from Google for a period of time because it created *doorway pages*—pages full of keyword-rich copy created solely for the search engine

spiders and never for human viewing. To add insult to injury, Google engineer Matt Cutts publicly outed BMW on his blog. He made an example of BMW, and all of the SEO community became aware of the carmaker's indiscretions.

Search engines rely primarily on automated means for detecting spam, with some auxiliary assistance from paid evaluators, spam vigilantes, and even your competitors. Search engineers at Google, Yahoo!, and Microsoft write sophisticated algorithms to look for abnormalities in inbound and outbound linking, in sentence structure, in HTML coding, and so on.

As far as the search engines are concerned, SEO has an acceptable side and an unacceptable side. In general terms, all types of actions intended to boost a site's search engine ranking without improving the true value of a page can be considered spamming.

Each search engine has different published guidelines. Here is where you can find them:

- Google's Webmaster Guidelines are located at <http://www.google.com/webmasters/guidelines.html>.
- Yahoo! Search Content Quality Guidelines are located at <http://help.yahoo.com/help/us/ysearch/basics/basics-18.html>.
- Bing Guidelines for Successful Indexing are located at http://help.live.com/help.aspx?mkt=en-us&project=wl_webmasters.

Each search engine has varying degrees of tolerance levels for various SEO tactics. Anything that violates these guidelines, pollutes the search results with irrelevant or useless information, or would embarrass you if your Google AdWords or Yahoo! rep discovered it, is unsustainable and should be avoided.

There's a big difference between "search-engine-friendly" and crossing the line into spam territory. "Search-engine-friendly" can mean, for example, that the site is easily accessible to spiders, even if it is database-driven; that HTML code is streamlined to minimize the amount of superfluous code; that important headings, such as product names, are set apart from the rest of the text (e.g., with H1 tags) and contain relevant keywords; or that link text is contextual, instead of comprising just "click here" or "more info" references.

Contrast these basic SEO practices with the following manipulative search engine spam tactics:

- Serving pages to the search engines that are useless, incomprehensible, unsuitable for human viewing, or otherwise devoid of valuable content—such as doorway pages, which SEO vendors may refer to by more innocuous names, including "gateway pages," "bridge pages," "jump pages," "attraction pages," "advertising pages," "channel pages," "directory information pages," "search engine entry pages," "satellite sites," "mini sites," "magnet sites," or "shadow domains." Whatever you call them, by definition they are created for the sole purpose of boosting search engine ranking.
- Creating sites with little useful unique content. There are many techniques for doing this, including:

- Duplicating pages with minimal or no changes and exposing them to the same search engines under new URLs or domains.
- Machine-generating content to chosen keyword densities (e.g., using Markov chains; you can read more about these techniques at <http://en.kerouac3001.com/markov-chains-spam-that-search-engines-like-pt-1-5.htm>, but they are not recommended).
- Incorporating keyword-rich but nonsensical gibberish (also known as *spamglish*) into site content.
- Creating a low-value site solely for affiliate marketing purposes (see the upcoming subsection on duplicate content for a more complete definition of *thin affiliate*).
- Repeating the same keyword phrase in the title tag, the H1 tag, the first alt tag on the page, the meta description, the first sentence of body copy, and the anchor text in links pointing to the page.
- Targeting obviously irrelevant keywords.
- Concealing or obscuring keyword-rich text or links within the HTML of a page so that it is not visible or accessible by human users (within comment tags, noscript tags, noframe tags, colored text on a similarly colored background, tiny font sizes, layers, or links that don't show as links to users because they are not highlighted in some manner, such as with an underline).
- Hijacking or stealing content from other sites and using it as fodder for search engines. This is a practice normally implemented using scrapers. Related to this is a practice known as *splogging*: creating blogs and posting stolen or machine-generated content to them.
- Purchasing links for the purpose of influencing search rankings.
- Participating in *link farms* (which can be distinguished from directories in that they are less organized and have more links per page) or *reciprocal linking schemes* (link exchanges) with irrelevant sites for the purpose of artificially boosting your site's importance.
- Pepperizing websites' guest books, blogs, or forums in bulk with keyword-rich text links for the purpose of artificially boosting your site's importance.
- Conducting *sneaky redirects* (immediately redirecting searchers entering your site from a keyword-rich page that ranks in the search engine to some other page that would not rank as well).
- *Cloaking*, or detecting search engine spiders when they visit and modifying the page content specifically for the spiders to improve rankings.
- Buying expired domains with high PageRank, or snapping up domain names when they expire with the hope of laying claim to the previous site's inbound links.
- *Google-bowling*, or submitting your competitors to link farms and so on so that they will be penalized.

These tactics are questionable in terms of effectiveness and dubious in the eyes of the search engines, often resulting in being penalized by or banned from the search engines—a risk that's

only going to increase as engines become more aggressive and sophisticated at subverting and removing offenders from their indexes. We do not advocate implementing these tactics to those interested in achieving the long-term benefits of SEO.

The search engines detect these tactics, not just by automated means through sophisticated spam-catching algorithms, but also through spam reports submitted by searchers—and yes, by your competitors. Speaking of which, you too can turn in search engine spammers, using the forms located at the following URLs:

- Google: <http://www.google.com/contact/spamreport.html>
- Yahoo!: http://add.yahoo.com/fast/help/us/ysearch/cgi_reportsearchspam
- Bing: <http://support.msn.com/feedbacksearch.aspx>

A lot of times marketers don't even know they're in the wrong. For example, some years ago JC Penney, engaged an SEO vendor that used the undesirable tactic of doorway pages. Consequently, a source at Yahoo! confirmed, unbeknownst to JC Penney its entire online catalog with the exception of the home page was banned by Yahoo! for many months. Over time, this must have cost the merchant a small fortune.

Marketers can even get caught in the crossfire without necessarily doing anything wrong. For example, search engines much more heavily scrutinize pages that show signs of potential deception, such as no-archive tags, noscript tags, noframe tags, and cloaking—even though all of these can be used ethically.

There is a popular myth that SEO is an ongoing chess game between SEO practitioners and the search engines. One moves, the other changes the rules or the algorithm, then the next move is made with the new rules in mind, and so on. Supposedly if you don't partake in this continual progression of tactic versus tactic, you will not get the rankings lift you want.

For ethical SEO professionals, this is patently untrue. Search engines evolve their algorithms to thwart spammers. If you achieve high rankings through SEO tactics within the search engines' guidelines, you're likely to achieve sustainable results.

Seeing SEO as a chess game between yourself and the search engines is often a shortsighted view. Search engines want to provide relevant search results to their users. Trying to fool the search engines and take unfair advantage using parlor tricks isn't a sustainable approach for anybody—the company, its SEO vendor, the search engine, or the search engine's users. It is true, however, that tactics, once legitimately employed, can become less effective, but this is usually due to an increased number of companies using the same tactic.

Recognizing Low-Quality Domains and Spam Sites

You can spot a poor-quality site in many ways. Search engines rely on a wide range of signals as indicators of quality. Some of the obvious signals are site owners who are actively spamming the search engines—for example, if the site is actively buying links and is discovered.

However, there are also less obvious signals. Many such signals mean nothing by themselves and gain significance only when they are combined with a variety of other signals. When a number of these factors appear in combination on a site, the likelihood of it being seen as a low-quality or spam site increases.

Here is a long list of some of these types of signals:

- Short registration period (one year, maybe two)
- High ratio of ad blocks to content
- JavaScript redirects from initial landing pages
- Use of common, high-commercial-value spam keywords such as *mortgage, poker, texas hold 'em, porn, student credit cards*, and related terms
- Many links to other low-quality spam sites
- Few links to high-quality, trusted sites
- High keyword frequencies and keyword densities
- Small amounts of unique content
- Very few direct visits
- Registered to people/entities previously associated with untrusted sites
- Not frequently registered with services such as Yahoo! Site Explorer, Google Webmaster Central, or Bing Webmaster Tools
- Rarely have short, high-value domain names
- Often contain many keyword-stuffed subdomains
- More likely to have longer domain names
- More likely to contain multiple hyphens in the domain name
- Less likely to have links from trusted sources
- Less likely to have SSL security certificates
- Less likely to be in directories such as DMOZ, Yahoo!, Librarian's Internet Index, and so forth
- Unlikely to have any significant quantity of branded searches
- Unlikely to be bookmarked in services such as My Yahoo!, Delicious, [Faves.com](#), and so forth
- Unlikely to get featured in social voting sites such as Digg, Reddit, Yahoo! Buzz, StumbleUpon, and so forth
- Unlikely to have channels on YouTube, communities on Facebook, or links from Wikipedia
- Unlikely to be mentioned on major news sites (either with or without link attribution)
- Unlikely to register with Google/Yahoo!/MSN Local Services

- Unlikely to have a legitimate physical address/phone number on the website
- Likely to have the domain associated with emails on blacklists
- Often contain a large number of snippets of “duplicate” content found elsewhere on the Web
- Unlikely to contain unique content in the form of PDFs, PPTs, XLSs, DOCs, and so forth
- Frequently feature commercially focused content
- Many levels of links away from highly trusted websites
- Rarely contain privacy policy and copyright notice pages
- Rarely listed in the Better Business Bureau’s Online Directory
- Rarely contain high-grade-level text content (as measured by metrics such as the Flesch-Kincaid Reading Level)
- Rarely have small snippets of text quoted on other websites and pages
- Cloaking based on user-agent or IP address is common
- Rarely contain paid analytics tracking software
- Rarely have online or offline marketing campaigns
- Rarely have affiliate link programs pointing to them
- Less likely to have .com or .org extensions; more likely to use .info, .cc, .us, and other cheap, easily obtained top-level domains (TLDs)
- Almost never have .mil, .edu, or .gov extensions
- Rarely have links from domains with .edu or .gov extensions
- Almost never have links from domains with .mil extensions
- Likely to have links to a significant portion of the sites and pages that link to them
- Extremely unlikely to be mentioned or linked to in scientific research papers
- Unlikely to use expensive web technologies (Microsoft Server and coding products that require a licensing fee)
- Likely to be registered by parties who own a very large number of domains
- More likely to contain malware, viruses, or spyware (or any automated downloads)
- Likely to have privacy protection on the Whois information for their domain

Some other signals would require data from a web analytics tool (which Google may be able to obtain from Google Analytics):

- Rarely receive high quantities of monthly visits
- Rarely have visits lasting longer than 30 seconds
- Rarely have visitors bookmarking their domains in the browser
- Unlikely to buy significant quantities of PPC ad traffic

- Rarely have banner ad media buys
- Unlikely to attract significant return traffic

For many, possibly even most, of these signals, there are legitimate reasons for a site to be doing something that would trigger the signal. Here are just a few examples:

- Not every site needs an SSL certificate.
- Businesses outside the United States will not be in the Better Business Bureau directory.
- The site may not be relevant to scientific research papers.
- The publisher may not be aware of Google Webmaster Tools or Bing Webmaster Tools.
- Very few people are eligible for an .edu, .gov, or .mil TLD.

These are just a few examples meant to illustrate that these signals all need to be put into proper context. If a site conducts e-commerce sales and does not have an SSL certificate, that makes that a stronger signal. If a site says it is a university but does not have an .edu TLD, that is also a stronger signal.

Many legitimate sites will have one or more of these signals associated with it. For example, there are many good sites with an .info TLD. Doing one, two, or three of these things is not normally going to be a problem. However, sites that do 10, 20, or more of them may well have a problem.

Competitors Can Report You

Search engines supplement their spam fighting by allowing users to submit spam reports. For example, Google provides a form for reporting spam at <http://www.google.com/contact/spamreport.html>. A poll at Search Engine Roundtable in May 2008 showed that 31% of respondents reported their competitors as spammers in Google. The bottom line is that having your competitor report you is a real risk.

In addition, the search engines can and do make use of human reviewers who conduct quality reviews. In fact, in 2007 a confidential Google document called the “Spam Recognition Guide for Raters” was leaked to the public (and it is still available at <http://www.searchbistro.com/spamguide.doc>). The guide delineates some of the criteria for recognizing search engine spam, such as whether the site is a thin affiliate.

Duplicate Content

As we discussed in “Content Management System (CMS) Issues” on page 263 in Chapter 6, there are many ways to create duplicate content. For the most part, this does not result from the activities of a spammer, but rather from idiosyncrasies of a website architecture. For this reason, the normal response of the search engines to duplicate content is to filter it out but not penalize the publisher for it in any other way.

They filter it out because they don't want to show multiple copies of the same piece of content in their search results, as this does not really bring any value to users. They don't punish the publisher because the great majority of these situations are unintentional.

However, there are three notable exceptions:

Copyright violations

An actual copyright violation where a publisher is showing a copy of another publisher's content without permission.

Thin affiliate sites

A scenario in which the publisher has permission from another publisher, but the content is not unique to them. The common scenario is a site running an affiliate network and generating leads or sales for their clients, largely by offering an affiliate program to other publishers.

The site running the affiliate program generates some content related to the offerings and distributes that to all their affiliates. Then all of these sites publish the exact same (or very similar) content on their websites. The problem the search engines have with this is that they offer very little value as there is nothing unique about their content.

This affiliate site may also create hundreds or thousands of pages to target vertical search terms with little change in content. The classic example of this is creating hundreds of web pages that are identical, except changing the city name referred to on each page (e.g., pages have titles such as "Phoenix Oil Changes," "Austin Oil Changes," "Orlando Oil Changes," etc.).

Massive duplication

Sites that have a very large amount of duplicate content, even if it is with permission, and it is not a thin affiliate. It is not known what the threshold is, and this probably changes over time, but our experience suggests that sites with 70% or more of their pages as duplicates of other pages on the Web are likely to be subject to a penalty.

There may be other spammy forms of duplicate content not identified here. It is likely to be spam if it is implemented intentionally (e.g., the thin affiliate site example qualifies under this metric), and the content adds no value to the Web.

Basic Rules for Spam-Free SEO

Especially if you are a novice SEO practitioner/publisher, the first and most important rule is to be familiar with the guidelines of the search engines (see the beginning of this section for the location of the guidelines on the search engine sites).

Second, it is essential that you learn to apply a basic personal filter to your SEO activities. Generally speaking, if you engage in the activity for the sole purpose of influencing search rankings, you are putting your site's rankings at risk.

For example, if you start buying keyword-rich text links from a bunch of sites across the Web and you do not expect to get significant traffic from these links (enough to justify placing the advertisement), you are headed for trouble.

Search engines do not want publishers/SEO experts to buy links for that purpose. In fact, Google has strongly campaigned against this practice. Google has also invested heavily in paid link detection. The following blog posts are from Matt Cutts, the head of Google's webspam team:

- <http://www.mattcutts.com/blog/text-links-and-pagerank/>
- <http://www.mattcutts.com/blog/quick-comment-on-nofollow/>
- <http://www.mattcutts.com/blog/selling-links-that-pass-pagerank/>
- <http://www.mattcutts.com/blog/text-link-follow-up/>

But there are more ways you can get into trouble other than purchasing keyword-rich text links. Earlier in this chapter we listed a number of ways that spammers behave. Most publishers/SEO practitioners won't run into the majority of these as they involve extremely manipulative behavior, as we outlined previously in this section. However, novice SEO practitioners do tend to make certain mistakes. Here are some of the more common ones:

- Stuffing keywords into your web page so it is unnaturally rich in those words.
- Over-optimizing internal links (links on your website to other pages on your website). Generally speaking, this may occur by over-stuffing keywords into the anchor text of those links.
- Cloaking, or showing different content to the search engine crawlers than you show to users.
- Creating websites with lots of very thin content pages, such as the thin affiliate sites we discussed previously.
- Implementing pages that have search-engine-readable text that is invisible to users (hidden text).
- Participating in link schemes, such as commenting on others' blogs or forums for the pure purpose of linking to yourself, participating in link farms, or using other tactics to artificially boost link popularity.

As we discussed previously, there are many other ways to end up in the crosshairs of the search engines, but most of those are the domain of highly manipulative SEO practitioners (sometimes referred to in the industry as *black hat SEO practitioners*).

Ultimately, you want to look at your intent in pursuing a particular SEO practice. Is it something you would have done if the search engines did not exist? Would it have been part of a publishing and promotional plan for your website in such a world?

This notion of intent is something that the search engines look at very closely when evaluating a site to see whether it is engaging in spamming. In fact, search engine representatives speak about *intent* (why you did it) and *extent* (how much did you do it) as being key things they evaluate when conducting a human review of a website.

A perceived high degree of intent and pursuing something to a significant extent is the realm of more severe penalties. Even if your intent is pure and you don't pursue a proscribed practice to a significant degree, the search engines will want to discount the possible benefits of such behavior from the rankings equation.

For example, if you buy a few dozen links and the search engines figure out that the links are paid for, they will discount those links. This may not be egregious enough for them to penalize you, but they don't want you to benefit from it either.

Identifying Search Engine Penalties

The reality is that penalties are imposed on websites. When this happens, it is useful to have an idea how to fix it. So, let's take a look at how it works.

The flowchart in [Figure 10-5](#) shows a series of basic checks that you can use to find out whether you have a problem. It is also important to remember that search engines are constantly tuning their algorithms (Google makes algorithm changes every day). Simple movement up and down in the rankings does not necessarily constitute a penalty.

Getting members of your management team who are not knowledgeable about SEO to understand this can be particularly challenging. Use this chart to help them better understand the difference between a penalty and natural movement in the rankings.

While you're at it, try to help your management team understand the landscape of search. Search engines make the rules, and since they provide the traffic, publishers have to follow them as best they can. This may not seem fair, but it is the way it works. If you can get a general commitment to follow the rules in your organization, you will have taken a big step forward.

Reinclusion/Reconsideration Requests

Once you have determined that you are subject to a penalty, there are two major steps you should take to restore your website to full search engine health.

The first step is to fix the problems that caused the penalty. This may not be easy to do if you have engaged in many practices considered manipulative by the search engines or if the violation of the guidelines you have been tagged for was accidental, but getting the penalty removed will not happen until you have fixed the problem.

Once the problems are fixed you can file a reinclusion/reconsideration request. However, we can't emphasize this enough: if you file such a request without first fixing the problem(s), if

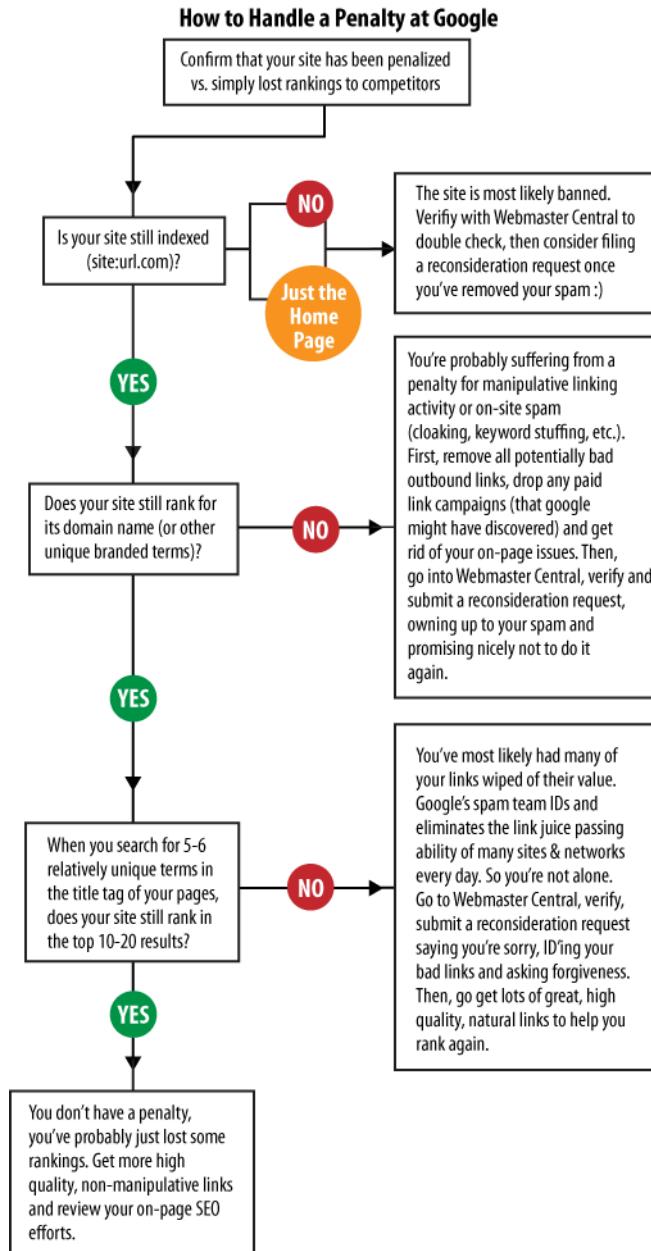


FIGURE 10-5. Determining whether you have a Google penalty

you were less than truthful, or if you omitted key facts in your request, you risk burning bridges with the engine and cementing the penalty.

First, your request may be ignored. Second, if you fix the problem and make the request again, that request may fall on deaf ears, in which case your site is more or less doomed.

So now, let's take a look at the process. The first step is to find the appropriate forms on the search engine's site. Here is where to find them:

Google

Go to Google Webmaster Tools (<https://www.google.com/webmasters/tools>), click on the Tools link, and then click the "Submit a reconsideration request" link.

Yahoo

The Yahoo! reinclusion request form (http://help.yahoo.com/l/us/yahoo/search/search_rereview_feedback.html) also provides a way to request a review of a penalty on your site. Fill out all the required fields and click Submit.

Bing

Go to Bing Webmaster Tools (<http://www.bing.com/webmaster>), go to the Bing Site Owner support form, fill in the required information, and indicate that this is a reinclusion request.

Here are the four basics of a reconsideration request, regardless of the engine:

- Understand the context of the situation with complete clarity. Search engines get a large number of these requests every day, and their point of view is that you violated their guidelines. You are asking for help from someone who has every right to be suspicious of you.
- Clearly define what you did wrong that led (to the best of your knowledge) to your site being banned or penalized.
- Demonstrate that you have corrected the issue(s).
- Provide assurances that it won't happen again. If you used an SEO firm that made the transgressions, make it clear that the firm is no longer working with you. If you did it yourself, be very contrite, and state that you now understand the right things to do.

Equally important is to have a clear idea of things you should not put into your reconsideration request:

- Do not use reconsideration request forms to request a diagnosis of why your site was banned. It will never happen. They get way too many of these per day.
- Don't whine about how much this is costing you. This is not material to their decision-making process, and it can irritate them. Remember that you need their goodwill.
- Do not compare their results with those of another search engine (in either traffic or rankings) to point out that something must be wrong. Search engine algorithms differ greatly, and the results you get will too.
- Don't mention how much you spend on advertising with that search engine. Search engines endeavor to keep their ranking teams separate from their advertising teams. Mentioning a large advertising spend will not help and can backfire if the request is interpreted as a request for special treatment.
- Do not submit a reconsideration request unless you are penalized, as we outlined previously.

The bottom line is to be succinct. Keep it short, crisp, and to the point. Research all of your current and past SEO activities to see which ones could be construed as an offense to the search engine's guidelines prior to any submission of a reconsideration request. You do not want to have the engineer look at the site and decide not to take action because you did not address every issue, because when you send in a second request, it will be that much harder to get their attention.

You should also be aware that you will probably never get a response, other than perhaps an automated one. If you have done your job well and have explained the situation well, you may just find your site back in the index one day.

Content Theft

The Web is sometimes a little bit of a Wild West environment, and one of the bad things that can occur is that another publisher (a spammer) can copy your content and republish it on one of their own sites. This is very easy for infringers to do. They simply "View Source" in their browsers and take whatever they like without regard to copyright, or send a spider to collect all of your pages in one fell swoop.

You can locate copyright infringers pretty easily using Copyscape (<http://www.copyscape.com>) if they've lifted some of your page copy. It's much more difficult if they've limited their sticky fingers to just your design. For example, Netconcepts had its design of [Netconcepts.com](#) stolen on more than one occasion, and the company discovered the issue only after a concerned visitor gave it a heads up.

There are six major options for dealing with an infringer:

- File a DMCA (Digital Millennium Copyright Act) infringement notification.
- Contact the infringing company's CEO, head of marketing, legal counsel, or whomever you feel is most likely to respond.
- Notify the site's hosting company, if it is not self-hosted.
- "Out" them on your blog, or if your blog is not that popular, get someone with a popular blog to out them.
- Have your lawyer send them a nasty-gram.
- Do nothing.

If you do nothing, eventually the infringer will redesign (probably stealing another design from somewhere else). That's not a great option if you're serious about protecting your intellectual property rights, however.

A phone call to the infringing company is inexpensive and it can work as well. But this isn't a viable option if you're dealing with an entity that operates overseas and has masked its domain contact information.

Luckily, the DMCA provides a useful hammer to beat on copyright infringers without the need to contact them, in the form of a DMCA infringement notification. In DMCA legal speak, this notification is also known as a "Takedown Notice."

As a content producer, you have the right to enforce your copyright. When your content gets "repurposed" on others' websites without your permission, you would simply file a DMCA Infringement Notification to the infringer's web hosting provider to get that infringer's website shut down (you can find a story about this on Ian McAnerin's site, <http://mcanerin.blogspot.com/2006/08/go-daddy.html>).

You can also get the infringer delisted from the search engines by filing DMCA notices to Google and the other major engines. It is not a daunting procedure. It might take an hour of your time, and it is well worth it. The process is as follows:

1. First, look up the web host and the domain registrar of the offending site using lookup tools such as Netcraft (<http://searchdns.netcraft.com/>) and DomainTools (<http://www.domaintools.com/>). You can usually ascertain who the web host is from the name servers and/or the netblock owner.
2. Check the official directory of designated DMCA agents (<http://www.copyright.gov/onlinesp/list/index.html>) for the host and the registrar. (Hopefully, they're listed.)
3. Prepare a letter to send to the designated agent of the web host. The notice you write should include your contact information, the name of the content that was copied, the web address of the copied content, a statement that you have a good faith belief that the material is not legal, a statement that under penalty of perjury you are the copyright

holder, and your signature. Some web hosts will allow you to email your notice to them, making it all that much more convenient.

4. Send a similar notification to the search engines. That will cut off their air supply (their traffic) in case the site doesn't get taken down right away. Instructions and contact details for each engine are available at the following URLs:

- Google: <http://www.google.com/dmca.html>
- Yahoo!: <http://info.yahoo.com/copyright/details.html>
- Bing: <http://www.microsoft.com/info/cpyrtInfrg.htm>

Note that Google requires you to mail or fax your letter, whereas Yahoo! and Microsoft (Bing) allow you to email your notification.

5. If the web host doesn't take the site down promptly, submit a DMCA notice to the infringer's domain registrar. It might be worth sending a notice to the data center that the web host uses before you try the registrar.

Be aware that the DMCA takedown procedure is a double-edged sword. Like any other tool, a DMCA notice can be used for good or evil. Specifically, someone could use the DMCA unfairly against you! It happens. Competitors do sometimes use the DMCA to silence their competitors (<http://it.toolbox.com/blogs/managing-infosec/using-dmca-to-silence-competitors-12745>).

Therefore, you as a website owner need to protect yourself from unwarranted (or at least unwelcome) prosecution. If the potential exists for you to inadvertently host infringing material on your website—for example, if you are hosting online forums, group blogs, blog comments, or other types of content that can be submitted from others besides yourself—here are some actions you can take to help protect yourself:

- It's helpful if you can qualify as a service provider (<http://www.chillingeffects.org/dmca512/faq.cgi#QID127>) that can be covered under the Safe Harbor provision (<http://www.chillingeffects.org/dmca512/faq.cgi#QID129>). For example, you may qualify if you offer a search engine or a bulletin board system.
- If so, notify your customers of your policies (<http://www.chillingeffects.org/dmca512/faq.cgi#QID128>) regarding copyright infringement and the consequences of repeated infringing activity. One way is to make it part of your Terms of Use.
- Also, publish a page on your website with DMCA filing instructions and state that, if and when you get a DMCA notification, you will act on it. An example of such DMCA notification instructions is available at <http://help.xanga.com/dmca.htm>.
- And most important, check the directory of designated agents; if your company isn't listed there, complete and file the required form (<http://www.copyright.gov/onlinesp/agent.pdf>) to the Copyright Office for inclusion in the directory.

Regardless of the path you choose, it is helpful to be able to simply and rapidly prove that you are the original author of the content. One tool that can help with that is the Wayback Machine

(<http://www.archive.org>). This site keeps copies of websites over time. You enter a website name, click Search, and then you might get a screen that looks a bit like Figure 10-6.



FIGURE 10-6. Sample data from archive.org

You can then click on one of the dates to see the state of the website on that date. The Wayback Machine does not always keep complete site copies, but a lot of data is available there, and you can use it to provide clear proof that you were the first person to publish a given piece of content.

Changing SEO Vendors or Staff Members

Another major transition that can happen is when the publisher changes its SEO staff. Adding new team members, people leaving the team, or replacing the whole team are all things that can and do happen. These types of changes can happen with in-house SEO staffs, and it can also happen when you use an external SEO agency (e.g., the publisher can decide to switch agencies, or the agency can have internal changes).

Companies make changes in their SEO staffs for many reasons, and issues can arise.

Potential Problems

One potential issue is a loss of momentum. The prior SEO team probably had some programs underway, and the new team will have to inherit those programs and sustain them, or else there will be a time delay in results from the new team's campaign.

Even if the new team does their best to sustain an existing campaign, it will take them some time to become familiar with it, and the specific approach the old team used may not fit the new team's skill set that well. It is best to be prepared for this type of loss of momentum when making significant changes in your SEO team.

One way to try to deal with this is to hire the new team and have some overlap with the old team. This can work in some situations, but it will not be an option in others.

Another situation that can occur is that the new team may have a different philosophy than the old team (this may be why the publisher made the switch). This will invariably mean a restart of efforts, potentially taking many weeks to regain the lost momentum.

Document SEO Actions and Progress

One way to limit the exposure to changes in the SEO team is through detailed documentation. Keep a detailed record of what has been done, when it was done, and why it is invaluable in these scenarios. Ideally, this should be combined with analytics data.

Traditional web analytics makes it easy to go back in time and see traffic data over time, but tools such as Google Webmaster Tools and Bing Webmaster Tools don't; nor do the search engines provide historical data on things such as indexed pages, link counts, and the like. You should also keep a record of these things over time.

In addition, make sure you have ownership of all relevant accounts and passwords. You don't want any transition to leave you in a situation where you can't access your critical data.

Rapid Training

Make sure that you have documented company processes regarding SEO, and that people associated with the SEO team and/or the management of the SEO team are familiar with those processes. This can help speed up transitions as well. The new team can be trained on what those processes are. Part of this includes having a detailed outline of the tasks and opportunities that are the first priorities for the new team.

With these things in hand, you can get things running smoothly much faster than you can without them. You can start the new team with one or more training sessions to get them up to speed on campaign history, login information, and what needs to be done going forward.

The better prepared you are to get them up to speed and moving forward, the better off you will be. As always, bear in mind that you have hired the new team for their expertise, so the initial training session may result in the creation of a modified plan, but you will get there much faster if you bring all these issues to the table on day one.

Clean Up

Last but not least, you should also go through and revoke the old team's access to your tracking systems. There is no need for them to continue to have access to your analytics, your Webmaster Tools accounts, and other resources that they no longer need.

Although the great majority of SEO firms (or individual SEO consultants) will handle the process of moving on with professionalism, there is no reason for you, as a publisher, to take that risk. So, although it might be exceedingly rare, the downside potential of leaving yourself vulnerable is just too great.

Conclusion

As they say, the only constant is change. Large-scale website changes are a fact of life, and they can seriously disrupt your rankings and traffic if not dealt with properly. Even when you do everything properly, it doesn't eliminate the risk of a sudden and significant drop in search traffic, but it certainly lowers your risk and recovery time.

CHAPTER ELEVEN

Honing the Craft: SEO Research and Study

LEARNING EVERYTHING YOU NEED TO KNOW ABOUT SEO IS A DIFFICULT PROCESS that is made even more challenging and complex by the constant evolution of best practices and optimization opportunities.

Fortunately, a great variety of resources that make the job easier are available on the Web. Many leading SEO practitioners regularly post their thoughts on blogs or in forums, or speak at conferences. This creates many opportunities to interact with them, learn from their experience, and keep your skills sharp and up-to-date.

In this chapter, we will talk about those resources and how to leverage them to be more successful in your SEO efforts.

SEO Research and Analysis

SEO is rapidly evolving. Search engines are constantly changing their algorithms, and new media and technologies are being introduced to the Web on a regular basis. Staying current requires an ongoing commitment to research.

SEO Resources

One of the easiest ways to research what is happening in the world of SEO is to study the websites and periodicals that cover SEO in detail, but ongoing testing of SEO hypotheses should also play a role.

Websites

A large number of online sites cover the search marketing space. Here is a short list of some of the most well-known ones:

- Search Engine Land (<http://www.searchengineland.com>), owned and operated by Third Door Media
- Search Engine Watch (<http://www.searchenginewatch.com>), owned and operated by Incisive Media
- SEOMoz (<http://www.seomoz.org>), owned and operated by SEOMoz

Each of these sites publishes columns on a daily basis, with Search Engine Land and Search Engine Watch publishing multiple posts every weekday. The columns are typically written by industry experts who have been chosen for their ability to communicate information of value to their reader bases. SEOMoz also provides a wide range of tools and resources for SEO practitioners.

Magazines

Magazines and journals that cover search and web marketing issues provide an additional information stream. Despite the fact that they provide less frequent coverage (as they are published monthly/bimonthly), the following are worth investigating:

- *Search Engine Marketing Research Journal* (<http://www.semj.org/>)
- *Search Marketing Standard* (<http://www.searchmarketingstandard.com/>)
- *Website Magazine* (<http://www.websitemagazine.com/>)

Commentary from search engine employees

Search engine representatives sometimes actively participate in forums, or publish blog posts and/or videos designed for webmasters. The main blogs for each of the three major search engines at the time of this writing are:

- Google Webmaster Central Blog (<http://googlewebmastercentral.blogspot.com/>)
- Bing's WebLog (<http://blogs.msdn.com/livesearch/>)
- Yahoo! Search Blog (<http://ysearchblog.com/>)

The search engines use these blogs to communicate official policy, announce new things, and provide webmasters with useful tips. You can reach Google personnel via the Google Webmaster Help group in Google Groups (http://groups.google.com/group/Google_Webmaster_Help/topics). Members of the Google webspam team are active in this group, answering questions and even starting their own new threads from time to time.

You can also interact with search engine representatives in various forums, such as WebmasterWorld (<http://www.webmasterworld.com>) and Search Engine Roundtable (<http://>

(forums.seroundtable.com/). Sometimes they use a nickname such as “googleguy” or “msndude,” so watch for those. You can also watch for search engine personnel who leave comments in popular SEO blogs. We will discuss the value of forums in more detail in “[The SEO Industry on the Web](#)” on page 500.

Interpreting commentary

Search engine reps are “managed” by their corporate communications departments. Some aren’t allowed to go on the record. Some need approval before doing so, and/or need their comments edited before publication. A rare few have free reign (e.g., Matt Cutts). Often they can’t be very specific or they can’t answer questions at all. The algorithms the search engines use are highly proprietary and they need to keep them secret.

This means there are certain types of questions they won’t answer, such as “What do I have to do to move from position 3 to position 1 on a particular search?” Or “How come this spammy site ranks so much higher than mine?”

In addition, they have their own motives and goals. They will want to reduce the amount of spam in their search index and on the Web overall (which is a good thing), but this may lead them to take positions on certain topics based on those goals.

As an example, Google does not talk about its capability for detecting paid links, but it suggests that its ability to detect them is greater than the general webmaster community believes. Taking this position is, in itself, a spam-fighting tactic, since it may scare people away from buying links who otherwise might have chosen to do so (as we indicated in [Chapter 7](#), we do not recommend purchasing links, but this example is meant to illustrate how a policy might affect communications).

In spite of these limitations, you can gather a lot of useful data from interacting with search engine representatives.

SEO Testing

SEO is both an art and a science. As with any scientific discipline, it requires rigorous testing of hypotheses. The results need to be reproducible, and you have to take an experimental approach so as not to modify too many variables at once. Otherwise, you will not be able to tell which changes were responsible for the results.

And although you can glean a tremendous amount of knowledge of SEO best practices, latest trends, and tactics from SEO blogs, forums, and e-books, it is hard to separate the wheat from the chaff and to know with any degree of certainty that an SEO-related claim will hold true. That’s where the testing of your SEO comes in: proving what works and what doesn’t.

Unlike multivariate testing for optimizing conversion rates, where many experiments can be run in parallel, SEO testing requires a serial approach. Everything must filter through the search engines before the impact can be gauged. This is made more difficult by the fact that

there's a lag between making changes and having the revised pages get spidered, as well as another lag while the spidered content makes it into the index and onto the search engine results pages (SERPs). On top of that, the results delivered depend on the user's search history, the Google data center accessed, and other variables that you cannot hold constant.

Sample experimental approach

Let's imagine you have a product page with a particular ranking in Google for a specific search term, and you want to improve the ranking and resultant traffic. Rather than applying a number of different SEO tactics at once, start varying things one at a time:

1. Tweak just the title tag and see what happens.
2. Continue making further revisions to the title tag in multiple iterations until your search engine results show that the tag truly is optimal.
3. Move on to the H1 tag, tweaking that and nothing else.
4. Watch what happens. Optimize it in multiple iterations.
5. Move on to the intro copy, then the breadcrumb navigation, and so on.

You can test many different elements in this scenario, such as:

- Title tag
- Headline (H1) tag
- Placement of the body copy in the HTML
- Presence of keywords in the body copy
- Keyword prominence
- Keyword repetitions
- Anchor text of internal links to that page
- Anchor text of inbound links to that page from sites over which you have influence

Testing should be iterative and ongoing, not just a "one-off" in which you give it your best shot and you're done. If you're testing title tags, continue trying different things to see what works best. Shorten it; lengthen it; move words around; substitute words with synonyms. If all else fails, you can always put it back to the way it was.

When doing iterative testing, it's good to do what you can to speed up the spidering and indexation so that you don't have to wait as long between iterations to see the impact.

You can do this by flowing more link juice to the pages you want to test. That means linking to them from higher in the site tree (e.g., from the home page). But be sure to give it some time before forming your baseline, because you will want the impact of changing the internal links to show in the search engines before initiating your test (to prevent the two changes from interacting).

Or you can use the Google Sitemaps protocol to set a priority for each page from 0 to 1.0. Dial up the priority to 1.0 to increase the frequency with which your test pages will be spidered.

N O T E

Don't make the mistake of setting *all* your pages to 1.0; if you do, none of your pages will be differentiated from each other in priority, and thus none will get preferential treatment from Googlebot.

Since geolocation and personalization mean that not everyone is seeing the same search results, you shouldn't rely on rankings as your only bellwether regarding what worked or didn't work.

Other useful SEO metrics

As we discussed in [Chapter 9](#), many other meaningful SEO metrics exist, including:

- Traffic to the page
- Spider activity
- Search terms driving traffic per page
- Number and percentage of pages yielding search traffic
- Searchers delivered per search term
- Ratio of brand to nonbrand search terms
- Unique pages spidered
- Unique pages indexed
- Ratio of pages spidered to pages indexed
- Conversion rate
- And many others

But just having better metrics isn't enough. An effective testing regimen also requires a platform that is conducive to performing rapid-fire iterative tests, in which each test can be associated with reporting based on these new metrics. Such a platform comes in very handy with experiments that are difficult to conduct under normal circumstances.

Testing a category name revision applied sitewide is harder than, say, testing a title tag revision applied to a single page. Specifically, consider a scenario where you're asked to make a business case for changing the category name "kitchen electrics" to a more search-engine-optimal "kitchen small appliances" or "small kitchen appliances". Conducting the test to quantify the value would require applying the change to every occurrence of "kitchen electrics" across the website. A tall order indeed, unless you can conduct the test as a simple search-and-replace operation, which you can do by applying it through a proxy server platform.

By acting as a middleman between the web server and the spider, a proxy server can facilitate useful tests that normally would be invasive on the e-commerce platform and time-intensive for the IT team to implement.

NOTE

During the proxying process, not only can words be replaced, but also HTML, site navigation, Flash, JavaScript, frames, even HTTP headers—almost anything. You also can do some worthwhile side-by-side comparison tests; a champion/challenger sort of model that compares the proxy site to the native website.

Start with a hypothesis

A sound experiment always starts with a hypothesis. For example, if a page isn't performing well in the engines and it's an important product category, you might hypothesize that this product category isn't performing well because it's not well linked from within your site. Or you may conclude that this page isn't ranking well because it is targeting unpopular keywords, or because it doesn't have enough copy.

Once you have your hypothesis, you can set up a test to gauge its truth. Try these steps:

1. In the case of the first hypothesis, link to that page from the home page and measure the impact.
2. Wait at least a few weeks for the impact of the test to be reflected in the rankings.
3. If the rankings don't improve, formulate another hypothesis and conduct another test.

Granted, this can be a slow process if you have to wait a month for the impact of each test to be revealed, but in SEO, patience is a virtue. Reacting too soon to changes you see in the SERPs can lead you to false conclusions. You need to give the search engines time to fully process what you have done so that you can improve the chances that you are drawing the right conclusions based on your tests. You also need to remember that the search engines may be making changes in their algorithms at the same time.

Analysis of Top-Ranking Sites and Pages

There are many reasons for wanting to analyze top-ranking sites, particularly those that are top-ranking in your market space. They may be your competitors' sites—which is reason enough to explore what they are doing—but even if they are not, it can be very helpful to understand the types of things they are doing to see how those things helped them get their top rankings. With this information in hand you can be better informed as you decide how to put together the strategy for your site.

Let's start by reviewing a number of metrics of interest and how to get them:

1. Start with a simple business analysis to see how a particular company's business overlaps with yours and with other top-ranked sites in your market space. It is good to know who is competing directly and who is competing only indirectly.
2. Find out the starting year for the website. This can be helpful in evaluating the site's momentum. Determining the domain age is easy; you can do it by checking the domain's Whois records. Obtaining the age of the site is trickier. However, you can use the Wayback Machine (<http://www.archive.org>) to get an idea of when a site was launched (or at least when it had enough exposure that [Archive.org](#) started tracking it).
3. Determine the number of Google results for a search for the domain name (including the extension) for the past six months, excluding the domain itself. To get this information, search for `theirdomain.com -site:theirdomain.com` in Google. Then append `as_qdr=m6` to the end of the results page URL and reload the page.
4. Determine the number of Google results for a search for the domain name (including the extension) for the past three months, excluding the domain itself. This time modify the results page URL by adding `&as_qdr=m3` to the end of it.
5. Perform a query on Google Blog Search (<http://blogsearch.google.com/>) for the domain name, excluding the domain itself, on the default settings (no particular timeline).
6. Find out from Google Blog Search how many posts have appeared about the site in the past month. To do this, search for the domain in Google Blog Search, but then append `&as_qdr=m1` to the end of the results page URL and reload the page.
7. Obtain the PageRank of the domain's home page as reported by the Google Toolbar.
8. Use Yahoo! Site Explorer to determine how many backlinks the site has. Better still is if you have an industrial-strength tool such as SEOMoz's Linkscape (<http://www.seomoz.org/linkscape>) or Majestic-SEO (<http://www.majesticseo.com>). These tools provide a much richer set of link data based on their own crawl of the Web, including additional critical details such as the anchor text of the links.
9. Pull a traffic chart from Compete (<http://www.compete.com>) that compares the site's traffic to that of its direct competitors to see how it is doing trafficwise. Repeat the traffic analysis with Quantcast (<http://www.quantcast.com>) and Alexa (<http://www.alexa.com>). Be careful to not put too much stock in the specific numbers, as the measurement techniques of these services are crude, but the relative traffic numbers should be information of great interest (e.g., competitor A is doing better than competitor B and how they both compare to you).
10. If you are able to access a paid service such as Hitwise (<http://www.hitwise.com>) or comScore (<http://www.comscore.com>), you can pull a rich set of additional data, breaking out the site's traffic by source (e.g., organic versus paid versus direct traffic versus other referrers). You can also pull information on their highest-volume search terms for both paid and organic search.

11. Determine the number of indexed pages in each of the three major search engines, using `site:theirdomain.com`.
12. If relevant, obtain Technorati's authority number for the site, which derives from the number of individual, unique blogs that have linked to a site in the past 90 days.
13. If relevant, get Google's feed subscriber numbers for the site, which you can find by searching for domains inside Google Reader.
14. If relevant, determine Bloglines' subscription numbers for the site, which derives from searches performed inside Bloglines.
15. Search on the company brand name at Google, restricted to the past six months (by appending `&as_qdr=m6` as outlined earlier).
16. Repeat the preceding step, but for only the past three months (using `&as_qdr=m3`).
17. Perform a Google Blog Search for the brand name using the default settings (no time frame).
18. Repeat the preceding step, but limit it to blog posts from the past month (using `&as_qdr=m1`).

Of course, this is a pretty extensive analysis to perform, but it's certainly worthwhile for the few sites which are the most important ones in your space. You might want to pick a subset for other related sites.

NOTE

As valuable as website metrics are, brand names can sometimes provide even more insight. After all, not everyone is going to use the domain name when talking about a particular brand, nor will they all link. Thus, looking at brand mentions over the past few months can provide valuable analysis.

Analysis of Algorithmic Differentiation Across Engines and Search Types

Each search engine makes use of its own proprietary algorithms to crawl and index the Web. Although many of the basic elements are the same (such as links being used as votes), there are significant differences among the different engines. Here are some examples of elements that can vary in on-page SEO analysis:

- Weight of title tags
- Weight of heading tags
- Emphasis on title tags
- Weight placed on synonyms
- Value of internal link anchor text
- How internal links are weighted as votes for a page

- Duplicate content filtering methods
- And many, many more

Similarly, there are many different ways a search engine can tune its algorithm for evaluating links:

- Percentage of a page's link juice that it can vote for other pages
- Weight of anchor text
- Weight of text near the anchor text
- Weight of overall linking page relevance
- Weight of overall relevance of the site with the linking page
- Factoring in placement of the link on the page
- Precise treatment of `NoFollow`
- Other reasons for discounting a link (obviously paid, manually tagged as paid, etc.)
- And many, many more

A detailed understanding of the specifics of a search engine's ranking system is not possible. However, with determination you can uncover various aspects of how the search engines differ. One tactic for researching search engine differences is to conduct some comparative searches across the engines. For example, if you search on *blog* in Google and Yahoo! you get the data shown in **Table 11-1**.

TABLE 11-1. Comparison of top five results for “blog” in Google and Yahoo!

Google	Yahoo!
http://www.blogger.com	http://www.blogger.com
http://googleblog.blogspot.com	http://360.yahoo.com
http://en.wikipedia.org/wiki/Blog	http://blogsearch.google.com
http://sethgodin.typepad.com	http://en.wikipedia.org/wiki/Blog
http://kanyeuniversecity.com/blog	http://wordpress.com

There are some pretty significant differences. For example, notice how Yahoo! has Google Blog Search as its #3 result. Google has most likely filtered this out of the results because it is a search engine home page, and Yahoo! has not. Yahoo! also has [360.Yahoo.com](http://360.yahoo.com) in the #2 position, whereas Google does not list it as a result at all.

You can also try to conduct some detailed analysis to guess why Seth Godin shows up in Google but not in Yahoo!, or why Yahoo! is showing the WordPress site. You may find that Google is weighting anchor text more heavily, whereas Yahoo! places greater value on contextual analysis.

It is also interesting to analyze the similarities. Both search engines have [Blogger.com](#) in the #1 position and the Wikipedia page in a high position. This speaks to things in common in the algorithms.

Using Experience and Instinct

There are some commonly perceived differences among the search engines. For example, Yahoo! is perceived to be more focused on contextual analysis than Google. Google is believed to place greater weight on link analysis than Yahoo! or Bing (note that link analysis *is* very important to all three engines).

There are also institutional biases to consider. Yahoo!, for example, has some of the world's most popular websites (e.g., [Yahoo.com](#)), and this gives Yahoo! access to data that the other search engines don't have from those sites. Given its huge market share, Google has the richest array of actual search data. Similarly, Microsoft has substantial assets in terms of its software and operating businesses, MSN (<http://www.msn.com>), and products such as Hotmail. The nature of the data that they have available to them can be influencing factors in how the search engines make their decisions.

Over time and with experience, you can develop a sixth sense for the SERPs so that when you look at a set of search results you will have a good grasp of the key factors in play without having to analyze dozens of them.

Competitive Analysis

Everything we discussed previously regarding analyzing top sites applies to analyzing competitors as well, and there are additional analysis methods that can help you gain a thorough understanding of how your competitors in search are implementing their SEO strategies.

Content Analysis

When examining a competing website, ask yourself the following questions:

- What content do they currently have on their site(s)? Answering this question will tell you a lot of things, including how they view their customers and how they use content to get links. For example, they may have a series of how-to articles related to their products, or a blog, or some nifty free tools that users may like. If they do have a blog, develop a sense as to what they write about. Also, see whether the content they are developing is noncommercial or simply a thinly disguised ad.
- How rapidly is that content changing? Publishers with rapidly changing websites are actively investing in their websites, whereas those who are not adding new articles or updating content may not be actively investing in their sites.

- What type of content is it? Articles? Videos? Images? Music? News feeds? The type of content they focus on can give you insight into their marketing strategy.
- Are they collecting user-generated content? Sites that gather a meaningful amount of user-generated content tend to have an engaged user audience.
- Are they trying to generate sign-ups or conversions in a direct way with their content? Or is it editorial in tone and structure? This can give you more visibility into the way their strategy is put together.

Internal Link Structure and Site Architecture

Your competitors' site organization and internal linking structure can indicate their priorities. Content linked to from the home page is typically important. For example, the great majority of websites have a hierarchy in which the major subsections of the site are linked to from the home page, and perhaps also from global navigation that appears on all or most of the pages on the site.

But what else is linked to from the home page? If the competitor is SEO-savvy this could be a clue to something that they are focusing on. Alternatively, they may have discovered that traffic to a given piece of content converts well into business. Either way, this is interesting to know.

External Link Attraction Analysis

You can extract a tremendous amount of information via link analysis. Yahoo! Site Explorer provides a start, but going past a simple list of links to gathering link data such as anchor text and link authority requires a more advanced tool, such as Link Diagnosis (<http://www.linkdiagnosis.com>), Linkscape (<http://www.seomoz.org/linkscape>), or Majestic-SEO (<http://www.majesticseo.com>).

In addition, you can do many other things by conducting an external link analysis of your competitors:

Dig deeper

What pages on the site are attracting the most links? These are probably the most important pages on the site, particularly if those pages also rank well for competitive search terms.

Determine where the content is getting its links

This can help you develop your own content and link-building strategies.

Analyze anchor text

Do they have an unusually large number of people linking to them using highly optimized anchor text? If so, this could be a clue that they are buying links.

If your analysis shows that they are not buying links, take a deeper look to see how they are getting optimized anchor text. This may be an indicator of a good strategy for you to use with your own site.

Determine whether they build manual links

Manual links come in many forms, so look to see whether they appear in lots of directories, or whether they are exchanging a lot of links.

Determine whether they are using a direct or indirect approach

Try to see whether they are using indirect approaches such as PR campaigns or social media campaigns to build links. Some sites do well building their businesses just through basic PR. If there are no signs that they are doing a significant amount of PR or social-media-based link building, it could mean they are aggressively reaching out to potential linkers by contacting them directly.

If they are using social media campaigns, try to figure out what content is working for them and what content is not.

Determine whether they are engaging in incentive-based link building

Are they offering incentives in return for links, such as award programs or membership badges? Do those programs appear to be successful?

Determine their overall link-building focus

Break down the data some more and see whether you can determine where they are focusing their link-building efforts. Are they promoting viral videos or implementing a news feed and reaching influencers through Google News and Yahoo! News?

What Is Their SEO Strategy?

Wrap up your competitive analysis by figuring out what their SEO strategy is. First, are they SEO-savvy? Here are some quick signals you can get that flag a competitor who is not SEO-savvy:

- Poorly optimized anchor text (if you see lots of links that say “Read more” or “Click here” it is a dead giveaway that they are not SEO-savvy).
- Content buried many levels deep. As you know from [Chapter 6](#), flat site architectures are a must in SEO.
- No signs of an active link-building campaign. Some sites can still rise to the top just because of a very effective (traditional) PR campaign.
- Critical content made inaccessible to spiders, such as content behind forms or content that can be reached only through JavaScript.
- And so forth.

Basically, if you see people making obvious errors and not following best or even good practices in basic ways, they are not very SEO-savvy. These competitors may still be dangerous if they

have a very strong brand, so you can't ignore them. In addition, they can fix a lack of SEO savvy by simply hiring a smart SEO expert.

Of course, if you are looking at a competitor who is SEO-savvy, read the clues on their site to figure out what they are doing.

Competitive Analysis Summary

Once you have all this information, you have to decide what to do with it. If your competition is implementing a strategy and it is working for them, you may want to follow it. However, that would potentially put you in a position of being a follower, so you may choose to focus on an area that appears to be a weak spot in their strategy.

For example, you might find that they are not paying attention to a segment of the market that has lots of potential linkers. Since you can stake out that portion of the market before they do, this can give you a competitive edge.

In addition, if they are buying links, what do you do? Do you turn them in using Google's paid links reporting form? Or do you copy them and buy links yourself? We do not recommend that you buy links, but if your competitor is doing it you do have some decisions to make.

Ultimately, you need to make some decisions about your own SEO strategy, and having a detailed understanding of what your competitors are doing can be invaluable in making those decisions.

Using Competitive Link Analysis Tools

It can be frustrating to see a competitor shoot ahead in the rankings and have no idea how they've achieved their success. Tools such as Linkscape, Link Diagnosis, and Majestic-SEO enable you to reverse the tactics that have produced valuable links and good anchor text by letting you quickly sort through the links pointing to a given site/page.

For example, if you were curious about how [Evosales.com](#) (a site selling scooters and mopeds) had earned its links, you could peek inside an advanced Linkscape report and see 852 links to the domain from 277 domains. Linkscape will show only the top 10 from any given domain, so even though it has 1,315 links, it's limiting some of those that come from domains with tons and tons of links (see [Figure 11-1](#)).

Looking at the data in [Figure 11-1](#), you can see a lot of links that appear to come from partnerships the company has with other scooter dealers on the Web, such as [Mopeds.com](#), [Neoscooters.com](#), and [Iscooterparts.com](#). Some nice natural links are coming from scooter forums such as [ScooterBBS.com](#), [Digg.com](#), [Kaboodle.com](#), and [AdLandPro.com](#). The company also has a small directory-link-building campaign earning some juice from sites such as [About-Cars.com](#), [rkom.com](#), [Motorcycleonlinestore.com](#), and more.

All in all, for a small number of links, the profile is pretty balanced and relatively organic.

The screenshot shows the Linkscape BETA interface. At the top, it says "Linkscape TM BETA" and "Linkscape provides link data never available before on over 54+ billion URLs across 230+ million domains." Below this, there's a search bar with "Enter a URL to Analyze" and a field containing "http:// www.dannydover.com". A "Run Basic Report" button is next to it. Below the search bar is a "+ Add URL" button. The main title is "Advanced Link Intelligence Report" for "www.dannydover.com" (generated Today, 16:54). To the right are "BOOKMARK" and "COMPARE" buttons. A message at the top of the report area says "This Advanced Link Intelligence Report has been saved. View later at: http://www.seomoz.org/linkscape/intel/report/44874" with a "Dismiss" button. Below this are navigation tabs: Dashboard, Links to URL (12), **Links to Domain (39)**, URL Anchor Text, Domain Anchor Text, and Custom Reports. The report itself is titled "Showing 1 – 39 of the 39 most important links". It includes sorting options (Sort by: mozRank Passed, Descending) and filtering options (Link with image, Link from noscript, Relnofollow, Link off screen, Same IP address, Same IP block, Same Subdomain, Same Root Domain). There are also "Anchor text" and "Search" filters. The results list the first link: "<http://www.dannydover.com/portfolio.html>". Below the link, it says "TITLE: Portfolio | Danny Dover" and "Anchor Text: Home". There are also small icons for social sharing.

FIGURE 11-1. Finding out your competitor's link-building tactics

For many cases, knowledge such as this will help your organization understand what the competition is up to. If there are lots of spammy, low-quality links, you can potentially report these to the engines. If the competition has great viral content, you can attempt to mimic or outdo their efforts.

And if they've simply got a small, organic footprint, you can be more aggressive with directories, content strategies, and direct link requests to overcome their lead. It's always an excellent idea to be prepared, and the ability to sort and filter out NoFollows and internal links and see where good anchor text and link juice flows from makes this process much more accessible than with other, less granular or expansive tools such as Yahoo! Site Explorer or Exalead.

Competitive Analysis for Those with a Big Budget

In [Chapter 5](#), we introduced you to Hitwise, which runs in the tens of thousands of dollars. Now we'll introduce you to two even more expensive packages: Covario (<http://www.covario.com>) and Conductor (<http://www.conductor.com>). Both are powerful, enterprise-grade SEO assessment and competitive analysis solutions that grade your website in a number of areas and compare it side by side to your competitors and their scores.

Using Search-Engine-Supplied SEO Tools

All three major search engines make an active effort to communicate with webmasters and publishers and provide some very useful tools for SEO professionals.

Search Engine Webmaster Tools

Although we have referred to the Webmaster Tools provided by Google and Bing throughout the book, it is worth going into these in greater depth. Using Webmaster Tools is a great way to see how the search engines perceive your site.

First, let's clear up a popular misconception. Setting up and using a Google Webmaster Tools or Bing Webmaster Tools account provides no new information about your site to the search engines. That is not the purpose of these tools.

You can easily get comfortable with this when you learn how the tools are set up. You can create a Webmaster Tools account with either Google or Bing quite easily. An important part of creating these accounts is verifying your ownership of the site. To do that you have two options:

- Place a search-engine-supplied meta tag on the home page of your domain.
- Place a special file as specified by the search engine in the root directory of your web server.

Neither of these actions changes or enhances the search engine's ability to access information about your website.

The intent of these tools is to provide publishers with data on how the search engines view their sites. This is incredibly valuable data that publishers can use to diagnose site problems. We recommend that all publishers leverage both of these tools on all of their websites.

For the next few pages, we will take a look at both of these products.

Google Webmaster Tools

[Figure 11-2](http://www.google.com/webmasters/tools) shows the type of data you get just by looking at the opening screen once you log in to Google Webmaster Tools (<http://www.google.com/webmasters/tools>).

You can find valuable data in the “not found” report. You can access this by selecting Diagnostics→Crawl Errors→Not Found, as shown in [Figure 11-3](#).

You can see a complete list of URLs that Googlebot encountered somewhere on the Web where your web server returned a 404 error instead of a web page. As we outlined in [Chapter 7](#), sometimes this happens because another publisher attempting to link to your site does not implement the link correctly.

For that reason, one great way to pick up some links is to simply implement a 301 redirect from the bad URL to the correct URL. If you can't figure out the correct URL you can always

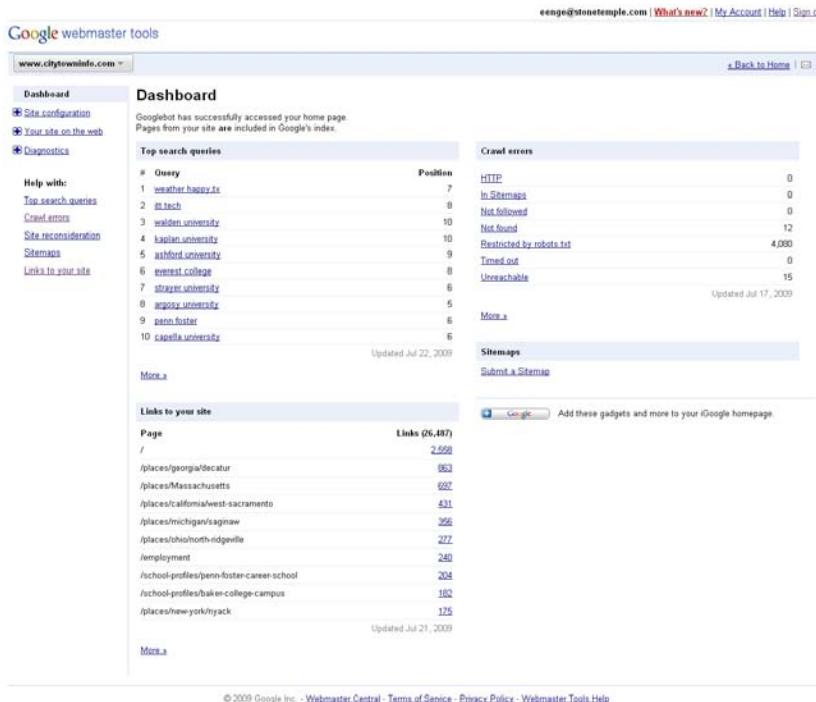


FIGURE 11-2. Google Webmaster Tools opening screen

redirect to the site's home page or the corresponding category one level up from the old URL. Before you do, though, check the pages that are linking to the broken pages, to make sure they represent legitimate links.

Another cool diagnostic tactic is to look at the error URLs in your Sitemaps file, as shown in [Figure 11-4](#).

With this data you can analyze the nature of the problem and resolve it. [Figure 11-5](#) shows another look at the diagnostic data.

The error shown in [Figure 11-5](#) is common, and occurs when sites mistakenly restrict access to URLs in their `robots.txt` file. When that happens, this report can be a godsend. The report flags pages it finds referenced on the Web, but that Google Webmaster Tools is not allowed to crawl.

Of course, this may be what you intended, in which case there is no problem. But when it is not what you intended this report provides you with some quick prompting to fix it.

Google Webmaster Tools also gives you an inside look at potential problems with your meta description tags and your title tags, as shown in [Figure 11-6](#).

Crawl errors

Issues Google encountered when crawling your site.

[Web](#) [Mobile CHTML](#) [Mobile WML/XHTML](#)

Show URLs: [HTTP \(0\)](#) - [In Sitemaps \(0\)](#) - [Not followed \(0\)](#) - [Not found \(14\)](#) - [Restricted by robots.txt \(0\)](#) - [Timed out \(0\)](#) - [Unreachable \(84\)](#)

URL	Detail	Linked From	Detected
http://www.stonetemple.com/STC_Background.shtml/	404 (Not found)	2 pages	Jul 13, 2009
http://www.stonetemple.com/STP_ResKevinSieck.htm	404 (Not found)	1 pages	Jul 9, 2009
http://www.stonetemple.com/articles/%20interview-rebecca-lieb.shtml	404 (Not found)	11 pages	Jun 30, 2009
http://www.stonetemple.com/articles/3..	404 (Not found)	unavailable	Jul 2, 2009
http://www.stonetemple.com/articles/analytics-report-august-2007%20.shtml	404 (Not found)	2 pages	Jul 1, 2009
http://www.stonetemple.com/articles/analytics-report-may-2007.shtml	404 (Not found)	7 pages	Jul 4, 2009
http://www.stonetemple.com/articles/interview-bruce-clay.shtml	404 (Not found)	1 pages	Jul 7, 2009
http://www.stonetemple.com/articles/interview-jakob-nielsen.shtml%20	404 (Not found)	3 pages	Jul 8, 2009
http://www.stonetemple.com/articles/interview-matt-cutts.shtml	404 (Not found)	3 pages	Jul 11, 2009
http://www.stonetemple.com/articles/interview-rand-schulman.shtml	404 (Not found)	2 pages	Jul 13, 2009
http://www.stonetemple.com/blog%20rel=external%20nofollow	404 (Not found)	16 pages	Jun 30, 2009

FIGURE 11-3. Google Webmaster Tools “not found” errors

Crawl errors

Issues Google encountered when crawling your site.

[Web](#) [Mobile CHTML](#) [Mobile WML/XHTML](#)

Show URLs: [HTTP \(0\)](#) - [In Sitemaps \(0\)](#) - [Not followed \(0\)](#) - [Not found \(3,805\)](#) - [Restricted by robots.txt \(2\)](#) - [Timed out \(0\)](#) - [Unreachable \(11\)](#)

URL	Detail	Detected
http://www.netconcepts.com/tag/e-commerce+website-audits+seo+business-blogging/	Failed to connect	Jul 21, 2009
http://www.netconcepts.com/tag/e-commerce+website-audits+seo+web-development/	Failed to connect	Jul 21, 2009
http://www.netconcepts.com/tag/search-optimization-video+seo+case-studies+blogs/	Failed to connect	Jul 21, 2009
http://www.netconcepts.com/tag/search-optimization-video+seo+case-studies+ecommerce/	Failed to connect	Jul 21, 2009
http://www.netconcepts.com/tag/social+business-blogging+seo+video+articles/	Failed to connect	Jul 21, 2009
http://www.netconcepts.com/tag/social+business-blogging+seo+video+examples/	Failed to connect	Jul 21, 2009
http://www.netconcepts.com/tag/web-dev-studies+ecommerce+link-building+blogs/	Failed to connect	Jul 21, 2009
http://www.netconcepts.com/tag/web-dev-se-studies+commerce+link-building+seo/	Failed to connect	Jul 21, 2009
http://www.netconcepts.com/tag/web-dev-copywriting+seminars+business-blogging/	Failed to connect	Jul 21, 2009
http://www.netconcepts.com/tag/web-dev-seo+copywriting+seminars+ecommerce/	Failed to connect	Jul 21, 2009
http://www.netconcepts.com/tag/web-marketing+business-blogging+online-marketing+seo/	Failed to connect	Jul 21, 2009

Updated Jul 24, 2009

[Download this table](#)

[Download all errors on this site](#)

FIGURE 11-4. Google Webmaster Tools “Sitemap errors” report

For example, if we dig a little deeper on the duplicate title tags we see the screen shown in Figure 11-7.

Figure 11-7 shows that there are eight cases of duplicate title tags on the site. Six of them are on the blog and two of them relate to a podcast with Neil Patel. These should be investigated to see whether there are problems that can be resolved.

Next, let’s look at the “top search queries” report, shown in Figure 11-8.

Crawl errors

Issues Google encountered when crawling your site:

Show URLs: [HTTP \(0\)](#) - [In Sitemaps \(0\)](#) - [Not followed \(0\)](#) - [Not found \(3,805\)](#) - [Restricted by robots.txt \(2\)](#) - [Timed out \(0\)](#) - [Unreachable \(11\)](#)

URL	Detail	Detected
http://www.netconcepts.com/redesign_your_email/	URL restricted by robots.txt	Jun 30, 2009
http://www.netconcepts.com/wp-content/plugins/wp-email/wp-email.php	URL restricted by robots.txt	Jul 20, 2009

[Download this table](#) | [Download all errors on this site](#)

FIGURE 11-5. Google Webmaster Tools URLs restricted by robots.txt

Google webmaster tools

www.citytowninfo.com | [eorge@stonetemple.com](#) | [What's new?](#) | [My Account](#) | [Help](#) | [Sign out](#)

[Back to Home](#) | [1](#)

HTML suggestions

When Googlebot crawled your site, it found some issues with your content. These issues won't prevent your site from appearing in Google search results, but addressing them may help your site's user experience and performance.

	Pages
Duplicate meta descriptions	217
Long meta descriptions	0
Short meta descriptions	468

Title tag

	Pages
Missing title tags	0
Duplicate title tags	1,204
Long title tags	0
Short title tags	0
Non-informative title tags	0

Non-indexable content

	Pages
We didn't detect any issues with non-indexable content on your site	0

[Download this table](#)

Last updated Jul 23, 2009

© 2009 Google Inc. - [Webmaster Central](#) - [Terms of Service](#) - [Privacy Policy](#) - [Webmaster Tools Help](#)

FIGURE 11-6. Google Webmaster Tools “metadata errors” report

Figure 11-8 shows Google’s view of the queries that are most searched on in terms of the site in question (in the left column), and the most frequently clicked queries for the site (in the right column). You can also see data from Google regarding what the site’s position was in the SERPs at that time. Note that this data is pretty limited, and most publishers will be able to get better data on their search queries from web analytics software.

Another report to look at is the “crawl stats” report, shown in Figure 11-9.

The charts in Figure 11-9 look normal and healthy for the site. However, if you see a sudden dip that sustains itself, it could be a flag that there is a problem.

Of course, one of the most important reports available from Google Webmaster Tools is the one that depicts the links to the site, as shown in Figure 11-10.

The screenshot shows the 'Google webmaster tools' interface for the domain 'www.stonetemple.com'. The left sidebar has a 'HTML suggestions' section selected. The main content area is titled 'HTML suggestions' and 'Duplicate title tags'. It explains that title tags provide information about the site and are likely to be clicked on. Below this is a table titled 'Pages with duplicate title tags'.

	Pages
# Ramblings About SEO > Blog Archive > 17 Poor Quality Signals Your Site May Be Sending	2
# Transcript of Podcast with Neil Patel	2
Download this table	

At the bottom, there's a copyright notice: © 2009 Google Inc. - [Webmaster Central](#) - [Terms of Service](#) - [Privacy Policy](#) - [Webmaster Tools Help](#).

FIGURE 11-7. Google Webmaster Tools duplicate title tags

The screenshot shows the 'Google webmaster tools' interface for the domain 'www.stonetemple.com'. The left sidebar has a 'Top search queries' section selected. The main content area is titled 'Top search queries' and provides information about user search queries. It includes two tables: 'Impressions' and 'Clickthrough'.

Impressions: Your site appeared in these searches

#	%	Query	Position	#	%	Query	Position
1	15%	page rank	12	1	24%	mark lucovsky	1
2	15%	page rank	12	2	6%	eric engr	1
3	10%	google custom search	7	3	4%	video search engine optimization	2
4	7%	mark lucovsky	3	4	4%	css and seo	1
5	5%	bounce rate	9	5	4%	css seo	1
6	4%	javascript location	9	6	4%	stone temple consulting	1
7	3%	osama	7	7	4%	netinsight	4
8	2%	local business center	10	8	3%	indie tools	2
9	1%	clicktracks	6	9	3%	stone temple	4
10	1%	google cse	9	10	2%	bad seo	1
11	1%	google custom	8	11	2%	cool things you can type	1
12	1%	stone temple	5	12	2%	eric jorgenson	1
13	1%	b2b	4	13	2%	grad conn microsoft	1
14	1%	amitash kaushik	13	14	2%	seo ssa	1
15	1%	rebecca leib	11	15	2%	b2b	3
16	1%	google custom search engine	8	16	1%	wordpress google description	1
17	<1%	custom search	9	17	1%	index tools	2
18	<1%	noscript tag	6	18	1%	shashi seth	2

Clickthrough: Users clicked on your site in these searches

#	%	Query	Position
1	mark lucovsky	1	
1	eric engr	1	
2	video search engine optimization	2	
1	css and seo	1	
1	css seo	1	
1	stone temple consulting	1	
4	netinsight	4	
2	indie tools	2	
4	stone temple	4	
1	bad seo	1	
1	cool things you can type	1	
1	eric jorgenson	1	
1	grad conn microsoft	1	
1	seo ssa	1	
3	b2b	3	
1	wordpress google description	1	
2	index tools	2	
2	shashi seth	2	

FIGURE 11-8. Google Webmaster Tools “top search queries” report

This report is the only way to get a good view of what links Google sees to the site. In addition, you can download the report in spreadsheet form, which makes it easy to manipulate the data. The final Google Webmaster Tools screen shot, shown in Figure 11-11, shows the Settings menu.

Here is what the various settings mean:

Geographic target

If a given site targets users in a particular location, webmasters can provide Google with information that will help determine how that site appears in the country-specific search results, as well as improve Google search results for geographic queries.

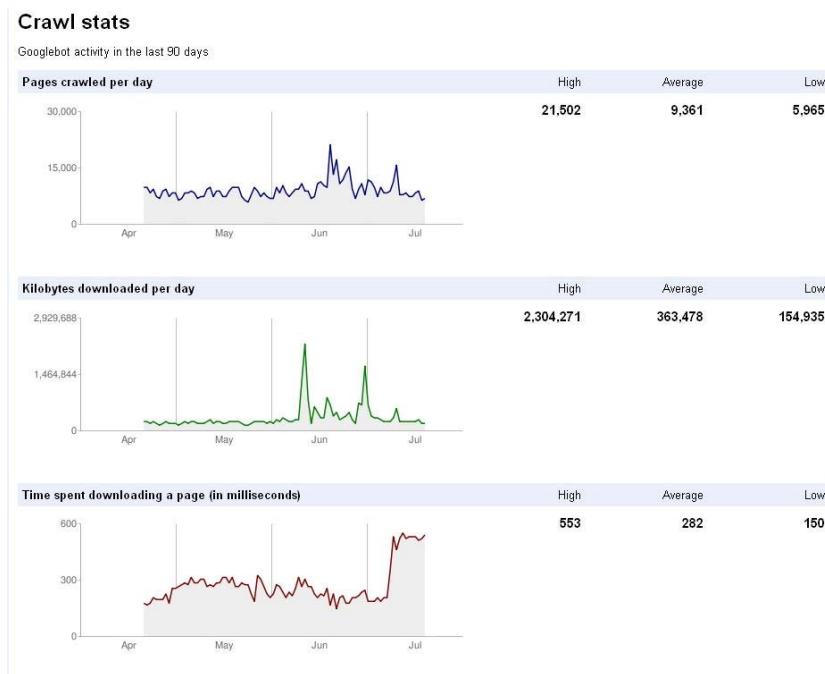


FIGURE 11-9. Google Webmaster Tools “crawl stats” report

Preferred domain

The preferred domain is the domain the webmaster wants to be used to index the site’s pages. If a webmaster specifies a preferred domain as <http://www.example.com> and Google finds a link to that site that is formatted as <http://example.com>, Google will treat that link as though it were pointing at <http://www.example.com>.

Crawl rate

The crawl rate affects the speed of Googlebot’s requests during the crawl process. It has no effect on how often Googlebot crawls a given site. Google determines the recommended rate based on the number of pages on a website.

robots.txt test tool

The *robots.txt* tool (not shown in Figure 11-11, but accessible at Site Configuration→Crawler Access→Test robots.txt) is extremely valuable as well. The tool allows you to test your *robots.txt* file using test URLs from your site before you actually push the file live.

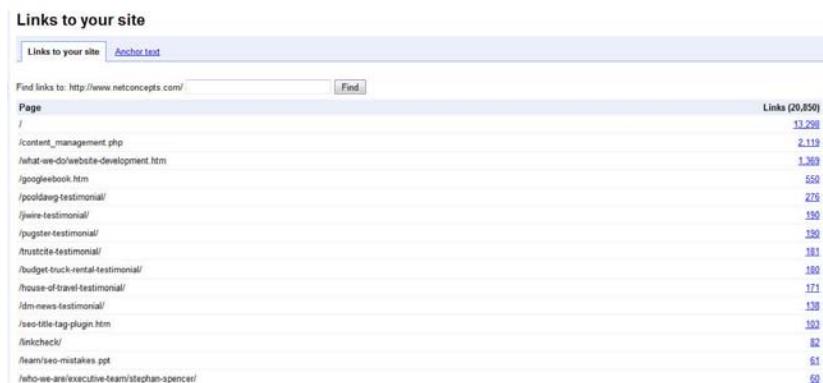


FIGURE 11-10. Google Webmaster Tools report depicting links to the site



FIGURE 11-11. Google Webmaster Tools Settings menu

Bing Webmaster Tools

Microsoft also offers a Webmaster Tools product with some great features. With Bing Webmaster Tools (<http://www.bing.com/webmaster>) you can get a rich set of data on many of the same types of things as Google Webmaster Tools offers, but now you are getting the feedback from the Bing crawler. The viewpoint from a different search engine is a tremendous additional value-add. Figure 11-12 shows the opening screen for the product.

Already you can see some of the great things that are available. At the top right you see the number of pages Bing has indexed. You can also see some of the most important pages, their page scores, and the most recent crawl date. Figure 11-13 shows the “crawl issues” report.

Site list

- Related tools
- Robots.txt validator
- HTTP verifier
- Keyword research tool
- Webmaster resources
- Forum
- Blog
- Help

Summary

Site status

Last crawled: 7/19/2009
Domain score:

Indexed pages: 2,890
Blocked: No

Top 5 pages

Page	Page score	Language	Region	Last crawled	Blocked
Strategic SEO Planning, Technical SEO and Link Building		en	us	7/16/2009	No
Ramblings About SEO » Blog Archive » Link Algorithm Changes		en	us	7/15/2009	No
Ramblings About SEO » Blog Archive » 11 Ways to Recognize a Bad...		en	us	7/19/2009	No
Ramblings About SEO » Blog Archive » Google Local Business Center...		en	us	7/14/2009	No
Ramblings About SEO		en	us	7/18/2009	No

© 2009 Microsoft | Privacy | Legal | Help | Feedback

FIGURE 11-12. Bing Webmaster Tools opening screen

Crawl issues

Select issue type: File Not Found (404)

Show all selected issues Filter selected issues by subdomain or subfolder [Learn more](#)

Include only results with this URL:

Search

Results for File Not Found (404)

Bing encountered a "404 File Not Found" HTTP status code when last attempting to download these URLs. This could either mean the page has been removed from your site, or that there is a misspelled URL somewhere on the web pointing to your site. We recommend you use this report to find a list of URLs to either repair the broken link, or create a 301 redirect the URL to a more appropriate location. [Learn more](#)

Results 1-3 of 3

Web Page	Date Last Found
1. http://www.stonetemple.com/thisisnotanexist.html	6/24/2009
2. http://www.stonetemple.com/scratching%20with%20a%20scratches	6/14/2009
3. http://www.stonetemple.com/oneleggermetaleas.com%20with%20a%20scratches	6/22/2009

Results 1-3 of 3

[Download all results](#) [Date Last Found](#)

© 2009 Microsoft | Privacy | Legal | Help | Feedback

FIGURE 11-13. Bing Webmaster Tools “crawl issues” report

You can pick from several options:

- The 404 report lists pages that Bing tried to access, but for which your web server returned a 404 status code.
- You can get a list of URLs blocked by *robots.txt*.

- You can find out where you have long dynamic URLs that may be troublesome to the Bing crawler.
- You can get a report of the pages on your site that appear to be infected with malware. Hopefully, you never need this, but if you do, it could be a lifesaver.
- You can get a review of all the URLs on your site that are of an unsupported content type.

Bing Webmaster Tools also provides you with a look at the set of inbound links that Bing is aware of for your site, as shown in [Figure 11-14](#).

The screenshot shows the Bing Webmaster Tools interface for the website www.stonetemple.com. The 'Backlinks' tab is selected. The results table displays 5 rows of backlinks, each with a page score (4 stars), language (en), and region (us). The columns are labeled: Web page, Page score, Language, and Region.

Web page	Page score	Language	Region
useit.com: Jakob Nielsen on Usability and Web Design	4	en	us
Link Building Expert Eric Ward's Linking Strategies, Linking Campaigns ...	4	en	us
Usability Effect Website Usability Reviews and Usability Consulting	4	en	us
An SEO & Web Marketing Blog: Scattereds	4	en	us
Strategic SEO Planning, Technical SEO and Link Building	4	en	us

FIGURE 11-14. Bing Webmaster Tools “inbound links” report

The backlink tool is limited in that you can download only 1,000 results, but it does provide a filtering capability so that you can dig into your backlinks in more detail, as shown in [Figure 11-15](#).

One unique feature is the ability to see which of your pages perform the best for a given search query within Bing, as shown in [Figure 11-16](#).

The screenshot shows the Bing Webmaster Tools interface for the site www.stonetemple.com. The left sidebar contains links for Site list, Related tools, Robots.txt validator, HTTP verifier, Keyword research tool, Webmaster resources Forum, Blog, and Help. The main content area has tabs for Summary, Profile, Crawl Issues, Backlinks (which is selected and highlighted in orange), Outbound Links, Keywords, and Sitemaps. The Backlinks section includes a search bar for filtering results by domain (e.g., edu) and a dropdown for 'Include only results with this URL'. Below this is a table titled 'Results 1-20 of 36' showing backlink details. The columns include 'Page score' (represented by a green bar), 'Language' (en), and 'Region' (us). The table lists five backlinks from the site, all of which are from the same source: <http://www.technet.microsoft.com>.

Page	Page score	Language	Region
Web Design References: Navigation	[Green Bar]	en	us
Heck with HAL: Open the Pod Bay Door to the Deep Web Yourself...	[Green Bar]	en	us
Welcome	[Green Bar]	en	us
Learning Remix Winter 2008: Tina Traitsas	[Green Bar]	en	us
Learning Remix Winter 2008: April 2008	[Green Bar]	en	us

FIGURE 11-15. Bing Webmaster Tools inbound link filtering

The screenshot shows the Bing Webmaster Tools interface for the site www.stonetemple.com. The left sidebar contains links for Site list, Related tools, Robots.txt validator, HTTP verifier, Keyword research tool, Webmaster resources Forum, Blog, and Help. The main content area has tabs for Summary, Profile, Crawl Issues, Backlinks, Outbound Links, Keywords (selected and highlighted in orange), and Sitemaps. The Keywords section includes a search bar for 'Keywords' (set to 'seo') and a 'Search' button. Below this is a table titled 'Web page' showing search results for the keyword 'seo'. The columns include 'Page score' (green bar), 'Language' (en), 'Region' (us), 'Last crawled' (date), and 'Blocked' (status). The table lists ten pages from the site, all of which have been crawled and are not blocked.

Page	Page score	Language	Region	Last crawled	Blocked
Strategic SEO Planning, Technical SEO and Link Building	[Green Bar]	en	us	7/26/2009	No
Ramblings About SEO	[Green Bar]	en	us	7/19/2009	No
Ramblings About SEO > Blog Archive > Five Strategic Steps for the ...	[Green Bar]	en	us	7/19/2009	No
Ramblings About SEO > Blog Archive > Free SEO Site Review	[Green Bar]	en	us	7/26/2009	No
Ramblings About SEO > Blog Archive > SEOs Must Work Closely With ...	[Green Bar]	en	us	7/25/2009	No
Ramblings About SEO > Blog Archive > "The most read site in the ...	[Green Bar]	en	us	7/19/2009	No
Ramblings About SEO > Blog Archive > Thoughts on SEO Ethics	[Green Bar]	en	us	7/25/2009	No
SEO Search from STC	[Green Bar]	en	us	7/19/2009	No
Ramblings About SEO > Blog Archive > Report Generating SEOs	[Green Bar]	en	us	7/25/2009	No
Ramblings About SEO > Blog Archive > 5 Principles for SEO Success	[Green Bar]	en	us	7/19/2009	No

FIGURE 11-16. Bing Webmaster Tools page performance by keyword

This type of data can be effective in helping you understand your best optimization strategy if you are trying to win on a given term. In Figure 11-16, Netconcepts probably should not focus on the “SEO: Metrics That Matter” page to try to rank for the query *SEO*, because so many other pages on the site are more highly ranked for that term.

However, there is an additional opportunity here. Since all the listed pages are relevant to the search query, it might be interesting to review which pages in this list link back to the top page in the list, and the anchor text used. After all, all of these pages have relevance to the query. This data can help you with your internal linking strategies.

There is much more here than we have been able to show in just a few pages. Bing Webmaster Tools provides you with the unique opportunity to see an insider's view of your website: to see it the way Bing sees it.

Yahoo! Site Explorer, Yahoo! Search Engine Link Commands

Yahoo! does not offer a Webmaster Tools product, but it does offer Yahoo! Site Explorer, and it is a very valuable tool indeed.

Within Yahoo! there are two ways to access link data: via Site Explorer, which provides a unique interface for browsing link data; and via the normal web search engine, which lets you apply many more advanced parameters and modifiers to your link-based queries.

Yahoo! Site Explorer

Yahoo! Site Explorer offers only a few basic features, including the ability to see:

- A list of links that point to a given URL
- A list of links that point to a given subdomain (e.g., [Southernfood.about.com](#), [Seomoz.org](#), or [Reddit.com](#))
- A list of links that point to a given root domain (e.g., *.about.com, *.seomoz.org, or *.reddit.com, including all their respective subdomains)
- A modifier that removes links coming from internal pages, either on the subdomain or the root domain, or from sister sites

However, the biggest feature is the ability to download up to 1,000 links into a spreadsheet file. In particular, you can download such a list for any domain at all—it does not have to be a site for which you have control, and in fact, it can be your competitor's website.

Using Site Explorer is simple. Just enter a given URL and go (or type a standard link or link domain search query into Yahoo! Search and you'll be redirected). You'll be given options to modify the domain parameters and exclusions in the results. [Figure 11-17](#) shows a sample report from Yahoo! Site Explorer.

Yahoo! Site Explorer's biggest weaknesses stem from the lack of crucial data pieces, including:

Ordering

The links are given in "no particular order," according to Yahoo!'s public representatives, though SEO practitioners generally believe they tend to show more important links before

The screenshot shows the Yahoo! Site Explorer interface for the domain nbcolympics.com. At the top, there's a search bar with the URL and a yellow "Explore URL" button. Below the search bar, the word "Results" is displayed. Underneath, there are two tabs: "Pages (160,418)" (which is selected) and "Inlinks (4,564)". To the right of these tabs are links to "Show pages from: All subdomains" and "Only this domain". Further down, there are "Result details:" dropdown menus and buttons for "Submit webpage or Site Feed" and "Export first 1000 results to TSV". The main content area displays a list of 7 results, each with a title, URL, file size, and cache status, followed by a "Explore" button. The first result is highlighted with a yellow background.

Rank	Page Title	URL	File Size	Cache Status	Action
1.	Vancouver 2010: NBCOlympics.com	text/html http://www.nbcolympics.com/	30k	- cache	Explore
2.	Olympic Video of Events, Competition, Athletes Olympics Highlights ...	text/html http://www.nbcolympics.com/video/index.html	23k	- cache	Explore
3.	2008 Beijing Summer Olympic Games TV & Online Listings Viewing ...	text/html http://www.nbcolympics.com/tv_and_online_listings	32k	- cache	Explore
4.	2008 Beijing Summer Olympic Games All-Time Medal Standings by Country ...	text/html http://www.nbcolympics.com/medals/index.html	102k	- cache	Explore
5.	2008 Beijing Summer Olympics - Beach Volleyball Athlete Photo ...	text/html http://www.nbcolympics.com/beachvolleyball/photos/galleryid=122861.html	38k	- cache	Explore
6.	U.S. Men Win 4x100m Freestyle Relay Video - NBCOlympics.com	text/html http://www.nbcolympics.com/video/share.html?videoid=0811_hd_swb_hd_I0194	52k	- cache	Explore
7.	2008 Beijing Summer Olympic Games Competition & Events Day-by-Day ...	text/html http://www.nbcolympics.com/resultsandschedules	141k	- cache	Explore

FIGURE 11-17. Yahoo! Site Explorer report sample

less important links (but certainly not in order from most to least valuable/popular/important/etc.).

NoFollow included

In a tragic move, NoFollowed links are included in the list with followed links, and no differentiation exists between the two, forcing SEO practitioners to do their own research page by page to determine which links the engines might actually be counting.

Target URL

Unless you choose to link to only a given URL, you don't get to see which page on a domain/subdomain a particular link points to.

Anchor text

No anchor text is provided to show the linking term/phrase/alt attribute that pointed to the page.

Importance metrics

No indication of how valuable/important a particular link or domain might be is provided. Obviously, Google's PageRank would be a strange one to show here, but Yahoo! used to have its own link-graph-based value, Webrank, which it quickly removed (likely only from public view) in 2004.

Despite these weaknesses, Yahoo! Site Explorer is at least valuable for browsing through a site's links and getting rough information on the types of sites and pages pointing to the URL/domain. If you want more, you can always click through to each individual link for some of this additional data (though this is a time-consuming process).

Yahoo! Search

Despite the fact that Site Explorer is meant to be the flagship link-searching product, Yahoo!'s normal query system actually provides far greater functionality. This includes the ability to get:

- A list of links to individual pages or entire domains (*link:http://www.yourdomain.tld* or *linkdomain:yourdomain.tld*)
- Lists that exclude pages from certain domains or top-level domains (*-site:domainx.com* or *-site:.co.uk*)
- Lists that include/exclude pages with certain keywords (*-keyword*, *-intitle:keyword*, *-inurl:keyword*, *keyword*, *intitle:keyword*, or *inurl:keyword*)
- Lists that include/exclude pages/domains with certain attributes (*-inurl:keyword* or *inurl:keyword*)
- Lists refined to certain domain extensions (*site:com.tr*)
- Lists refined by geographic region (*region:europe*)
- Lists of pages that contain links to other pages or domains (*linkdomain:domainx.com*, *linkdomain:domainy.com*)

Using Yahoo! Search to explore links requires modifying the standard link queries. If you simply use a query structure such as *linkdomain:nbcolympics.com* (to see links to a domain) or *link:http://www.nbcolympics.com* (to see links to an individual page), you'll be redirected to Site Explorer. You need to combine these queries with additional parameters to get data from Yahoo! Search, as shown in [Figure 11-18](#).



FIGURE 11-18. Yahoo! Search Linkdomain: command

The weaknesses here are the same as those for Site Explorer: no link targets, no anchor text, and no importance metrics. The only real advantage over Site Explorer is the ability to refine the query.

Here are some ways to use Yahoo!'s link data:

- Determine the approximate number and strength of a site's links (to help uncover the role that link strength/weakness is playing in rankings and search traffic).
- Track relative link growth/shrinkage over time.
- Uncover links that could be causing potential problems (paid links, spammed links, low-quality links, etc.) or drops in rankings (not necessarily because they made a site fall just by existing, but because their value might be removed).
- Find links that point to pages that return a 404 error.
- Determine the relative link popularity of a given page on a site compared to others.
- Investigate the links to a domain or page from which you're considering getting a link.
- Find links to competitors that outrank a client for potential acquisition.
- Find links that point with less-than-ideal anchor text to which you can request modifications.
- Research a site's internal link structure to find problems or opportunities.
- Check links between link partners or multiple sites owned by an entity.

The SEO Industry on the Web

SEO-specific blogs, forums, and social networking communities are almost always abuzz with new information and insights on the topic of SEO. Following this online community closely can help increase your SEO knowledge and keep your skills sharp.

Blogs

The SEO industry has a significant number of people who regularly publish high-quality content about SEO and web marketing. This provides a resource for newcomers who want to learn, and even for experienced SEO practitioners who can pick up tips from others.

The reality is that SEO is a deep and complex topic, and the opportunity for learning never stops. That is one of the most rewarding aspects of the profession. With the large number of blogs about SEO and web marketing in existence, it makes it easy to research the profession on a regular basis.

However, just about anyone can create a blog and start writing about SEO with self-appointed authority. There is a big difference between someone publishing something and what she publishes actually being true. Many of these people have been able to create consulting businesses out of "flash in the pan" service offerings rooted in opportunism, as opposed to experience and love of the mystery, dynamism, and challenge of search. However, as the economy shakes itself out to rebuild, so too will the search landscape. At all stages of your SEO

research process and lifelong learning curve, it is important to learn how to filter out those who know what they are talking about and those who do not.

One way to do that is to see which bloggers have the respect of other senior people in the industry. This is not unlike the process search engines go through to identify trusted authority sites and see who they recommend and trust. Proximity to clearly authoritative figures is a big plus.

Another way to do that is to become a contributor yourself. As you accumulate SEO experiences, share the ones you can. Comment in blogs (or forums) and become known to the community. Make sure you contribute to the community, and the community will reward you with help when you need it.

The search technology publications we discussed at the beginning of this chapter are a great place to start. However, a number of other SEO blogs provide a great deal of useful information. Here are some of the best, listed alphabetically.

First up are blogs by the authors:

Alchemist Media's The Alchemy of Search (<http://www.alchemistmedia.com/blog>)

Jessie Stricchiola's musings on the search industry

Ramblings About SEO (<http://www.stonetemple.com/blog>)

Eric Enge's thoughts on SEO and PPC, as well has his highly respected interviews of search engine representatives and industry insiders

SEOMozBlog (<http://www.seomoz.org/blog>)

Daily articles on issues of importance to the industry, and one of the most highly read blogs in the SEO space

Stephan Spencer's Scatterings (<http://www.stephanspencer.com/>)

The ruminations of Stephan Spencer, with detailed dives into many specific aspects of SEO

Here is a list of blogs by the search engines:

- Bing's WebLog (<http://www.bing.com/community/blogs/search/>)
- Google Webmaster Central Blog (<http://googlewebmastercentral.blogspot.com/>)
- Yahoo! Search Blog (<http://ysearchblog.com/>)

Although it is technically considered "unofficial," you should also make Matt Cutts's blog a must-read on your list. One of Google's search engineers and the head of Google's webspam team, Matt is also Google's official representative to the webmaster world.

- Matt Cutts's blog (<http://www.mattcutts.com/blog>)

Here are some other highly respected blogs:

Cre8pc (<http://cre8pc.com/>)

Kim Krause Berg on usability, marketing, website development, and the search markets

Dave Naylor's SEO Blog (<http://www.davidnaylor.co.uk/blog>)

Dave Naylor's coverage of a wide range of online marketing issues, with a focus on SEO
Google Blogoscoped (<http://blogoscoped.com/>)

Philipp Lenssen's journal of Google events, with an SEO bent
Graywolf's SEO Blog (<http://www.wolf-howl.com/>)

Michael Gray's thoughts on online marketing, with a few rants about Google mixed in
Marketing Pilgrim (<http://www.marketingpilgrim.com/>)

Andy Beal's journal of the search engine space and SEO events
Search Engine Journal (<http://www.searchenginejournal.com>)

Loren Baker's collection of posts and guest writers about events and phenomena in SEO and paid search

Search Engine Roundtable (<http://www.seroundtable.com>)
Barry Schwartz's roundup of all things search-related

SEOBook (<http://www.seobook.com>)
Aaron Wall's accompaniment to his excellent book on SEO

SEO Scoop (<http://www.seo-scoop.com>)
Donna Fontenot's personal and often introspective take on SEO

Small Business SEM (<http://www.smallbusinesssem.com/>)
Matt McGee's blog focused on search marketing for small businesses

StuntDubl (<http://www.stundubl.com>)
Todd Malicoat's SEO tips and tricks journal

You can also get a more comprehensive list by visiting Lee Odden's BIGLIST of SEO and SEM Blogs (<http://www.toprankblog.com/search-marketing-blogs/>).

Forums

Forums are a bit more free-form than blogs. People can create an account easily and say what they want. In a blogging environment, there are those who write about a topic and those who then comment on it. In a forum environment, most anyone can start a new thread or start to dominate an existing thread.

This makes the process of filtering even tougher than it is with a blog. People may have a strong opinion but be completely incorrect. As an example of this, you can still encounter people who believe that the primary task of an SEO practitioner is to decide on your meta tags. In 2001 this may have been effective and sufficient, but SEO has evolved a very long way since then.

That being said, there is a lot to be gleaned from the forums. Reading them does require a critical eye, but you can pick up a lot of value there. For example, tedster (Ted Ulle), who is a moderator at WebmasterWorld (<http://www.webmasterworld.com>), really knows his stuff and shares lots of useful information.

As with blogs, you want to see which people get the respect of their peers. Some forums rank or rate contributors based on the quantity of their contributions or votes. That is usually a good signal that they share valuable information. Even among people who are knowledgeable, not every statement should be accepted at face value. SEO is a very complex and deep topic, and is subject to lots of interpretation.

Once you develop your own feeling for SEO, though, this interpretation becomes increasingly easier. And as with the blogger community, there is value in being an active participant. Become a contributor, and when you have a question for the group, you are more likely to get useful answers.

You can also encounter search engine representatives on the forums. From time to time, each search engine has had one or more people who regularly look through the forums. The head of Google's webspam team, Matt Cutts, is known to respond to discussions in threads on the forums. Sometimes these responses contain interesting tidbits of information that provide a look inside the mindset of Google.

Here is an alphabetical list of some of the top SEO forums:

Cre8asite Forums (<http://www.cre8asiteforums.com/>)

Cre8asite is an open community of dedicated SEOs, usability professionals, web developers, and others who come together to provide some of the best advice related to online marketing and development on the Net. It's also where you'll find me (randfish) moderating and posting on a daily basis.

Digital Point Forums (<http://forums.digitalpoint.com/>)

This is Shawn Hogan's incredibly popular (and somewhat noisy) forum on everything related to the Web.

High Rankings Forums (<http://www.highrankings.com/forum/>)

Jill Whalen, a very well-known and respected SEO practitioner, runs High Rankings, a place where you'll also find experts such as Scottie Claiborne, Dan Thies, and Alan Perkins.

Search Engine Roundtable Forums (<http://forums.seroundtable.com/>)

These are the forums associated with Barry Schwartz's Search Engine Roundtable blog (<http://www.seroundtable.com>).

Search Engine Watch Forums (<http://forums.searchenginewatch.com/>)

This is where you'll find the Search Engine Watch website's online forums, a place where many high-visibility folks in the SEO/paid search space come to post.

WebmasterWorld Forums (<http://www.webmasterworld.com/>)

These are Brett Tabke's forums, which are incredibly popular and among the most popular places for SEOs to discuss issues.

WebProWorld Forums (<http://www.webproxworld.com/index.php?>)

This is a large web development and SEO forum with a very diverse group of posters.

In addition, the Search Engine Roundtable blog (at <http://www.seroundtable.com/>) provides a compendium of the most interesting notable forum threads with commentary/analysis added in. This serves as a useful overview if you can't keep up with all the forums individually.

Communities in Social Networks

A significant number of SEO practitioners are very active on social networking sites, such as:

- Facebook (<http://www.facebook.com>)

Our Facebook URLs are:

- Eric Enge: <http://www.facebook.com/people/Eric-Enge/565117597>
- Rand Fishkin: <http://www.facebook.com/people/Rand-Fishkin/501142015>
- Stephan Spencer: <http://www.facebook.com/people/Stephan-Spencer/519192602>
- Jessie Stricchiola: <http://www.facebook.com/profile.php?id=541298615>

- Twitter (<http://www.twitter.com>)

Our Twitter URLs are:

- Eric Enge: <http://twitter.com/stonetemple>
- Rand Fishkin: <http://twitter.com/randfish>
- Stephan Spencer: <http://twitter.com/sspencer>
- Jessie Stricchiola: <http://twitter.com/Itstricchi>

- LinkedIn (<http://www.linkedin.com>)
- Sphinn (<http://www.sphinn.com>)

Participating in social networks creates opportunities to interact with and learn from other SEO practitioners. As with the forums, you can also encounter search engine employees on these sites as well. Facebook, Twitter, and LinkedIn are all about social networking. None of these three sites is designed specifically for SEO practitioners, but you can meet well-known SEO folks through them, develop relationships with them, and participate in discussions about SEO topics.

Sphinn is run by Third Door Media (<http://thirddoormedia.com/>), and is a site that is designed specifically for SEO practitioners and covers the latest SEO news. As with Digg, users can submit articles to the site and these get voted on. When an article gets enough votes within a certain period of time it gets promoted to the home page of the site. This can be a great way to filter through the hundreds of posts about SEO that appear every day.

Participation in Conferences and Organizations

Conferences are another great way to meet leaders in the search space and network with them. The various panels may cover topics about which you are looking to learn more, and this by

itself is a great resource. Even experienced SEO practitioners can learn new things from the panel sessions.

Make sure you target the panels with the following objectives in mind:

- Attend panels in which you can pick up some information, tips, and/or tricks that you need for your immediate SEO projects or that broaden your SEO knowledge.
- Attend panels in which someone that you want to get to know is speaking.

N O T E

If you want to meet someone and she is speaking on a panel, sit in the front row, and at the end of the session introduce yourself and/or ask a question. This is an opportunity to get some free and meaningful advice from experts. Get the speaker's business card and follow up with her after the event.

Of course, you may encounter the people you want to meet in the hallways in and about the conference. Just go up and introduce yourself. You will find that most attendees are very open and will be happy to share information and tips.

N O T E

Some panels also have search engine representatives speaking on them, so this tactic can be useful in meeting them as well.

In addition to the day's sessions, each conference has a series of networking events during the day and sometimes in the evening. Many times these events are held in the bar of the hotel where (or near where) the conference is being held.

There is also a published party circuit, with different companies sponsoring parties every night of the conference (sometimes more than one per night). Consider going to these events as a way to meet people and build new relationships as well.

Just remember that the speakers deserve your respect. They are there to do business just as you are, and they may be able to answer questions you have or they may have other things on their minds. Be patient and treat them with respect, and this should provide you with the best results.

Here is an alphabetical list of some recommended conferences for search-specific content:

Ad-Tech

This is an interactive marketing conference that has begun to address many of the issues of SEO and search marketing.

Online Marketing Summit

Centered around Internet marketing of all kinds (but with at least one track devoted to SEO), OMS is a rising star in the conference world. One of the most valuable aspects comes from the one-on-one “lab” sessions where expert panelists sit down with attendees for 15- to 20-minute review sessions.

PubCon

Two to three times per year, Brett Tabke, the owner of WebmasterWorld, runs conferences that attract a large group of SEO practitioners.

Search Engine Strategies

Each year, SES runs 6 to 10 conferences worldwide. SES San Jose is the biggest event, but SES is also quite substantial.

Search Marketing Expo

SMX runs a number of conferences worldwide, including topically focused ones such as SMX Social Media and SMX Local. SMX also has the only search conference specifically for advanced-level SEO practitioners, SMX Advanced.

SEOmoz Training Seminars

Operated by SEOmoz, these seminars run annually in August in Seattle and in October in London, and focus on expert-level content and audiences.

In terms of more in-depth workshops, the association SEMPO (Search Engine Marketing Professionals Organization) offers a Certification program (<http://www.sempoinstitute.com/>), as does the **Direct Marketing Association** and **Market Motive**. The American Marketing Association offers a two-day SEO training intensive (<http://www.marketingpower.com/seo>) taught by Stephan Spencer.

Conclusion

Mastering SEO requires a great deal of study and effort. Even when you achieve initial mastery, you have to continue your studies to stay current. Thankfully, industry resources abound: blogs, forums, conferences, workshops, training materials, competitive analysis tools, and search-engine-supplied tool sets and knowledge centers. Take advantage of the rich sources of information and participate in the vibrant online community to stay in front of the learning curve, and your competition.

CHAPTER TWELVE

Build an In-House SEO Team, Outsource It, or Both?

The Dynamics and Challenges of Using In-House Talent Versus Outsourcing

ONCE YOUR ORGANIZATION HAS EMBRACED THE NEED FOR SEO, you must decide whether to hire someone in-house or work with an outside consultant or agency. The same question applies to expanding an existing SEO team, and you should make your decision based on the following:

- Is there enough work to justify a full-time, part-time, or contract employee?
- Is in-house knowledge on the topic strategically necessary for the business?
- Are people of sufficient skill available for hire within your target salary range?
- Do you want to access and leverage experienced talent from an external agency to supplement an internal team?

The Value of In-House SEO

One of the significant advantages of in-house SEO is the ability to access search expertise whenever you need it. Some additional advantages are:

- The greater likelihood that SEO best practices will be integrated into the company culture
- Greater accountability
- Faster ramp-up time (familiarity with the core business)
- Faster implementation time (greater leverage in the development process)

The Value of Outsourced Solutions

Although it may seem that in-house SEO is the ideal solution, there are also plenty of reasons to use an outsourced team instead of or in conjunction with an in-house team. Here is a summary of some of the best reasons to leverage outsourced talent:

- Finding expert-level (10+ years) SEO talent to hire is difficult, so it may be smarter to contract with such talent.
- It allows you to focus on your core business and outsource your noncore activities.
- Experts generally have a broader reach in the marketplace via industry contacts and as a result of working with many clients.
- SEO is multidisciplinary, so using external people makes it easier to pull in components when needed.
- Outsourced teams don't have the myopic focus that sometimes occurs in long-time in-house personnel.

Many SEO consulting firms, such as those of the authors, are willing to take on contracts that include training in-house personnel. In this environment, you may have the best of both worlds: the ability to immediately inject a high level of expertise into your organization (albeit on a temporary basis), and a long-term plan to bring the capability in-house, or at least build up the expertise with ongoing or occasional augmentation via outside consulting.

Leveraging SEO Knowledge in an Organization

Although it makes sense to involve an experienced SEO professional (employee or consultant) in your optimization efforts, it is important to remember that for SEO to be successful in the long term, you want SEO expertise to be involved in every aspect of your online business.

A basic level of SEO knowledge within the software and website development departments can speed up the production process, as a well-educated development team already knows a lot about how SEO affects what they do every day and can modify their efforts on an ongoing basis. They will be more likely to make the right decisions when the person in charge of SEO is not immediately available.

When you can address basic SEO knowledge within the development team, you can dramatically reduce the work for the SEO practitioner, so that when he does review your website for search engine friendliness and to create an SEO strategy, instead of spending time

talking to you about Flash, JavaScript, header tags, and alt text, he can focus on more advanced solutions and strategies.

Similarly, basic SEO knowledge is an incredibly powerful tool in both online and offline marketing. Many companies have failed to integrate their offline marketing efforts (such as television commercials promoting a new product) with their online efforts, and SEO planning in this area can often make (or break) a product or website launch.

SEO knowledge at the management level is essential for similar reasons. Decisions made by management, which can affect the entire business, can often have disastrous effects on SEO efforts if they are done in a vacuum. It is a common problem that even after the development teams, designers, and copywriters are trained on SEO, managers end up becoming the obstacle because they haven't been briefed on (or haven't bought into) the aspects of SEO that executives need to know.

When all of the members of an organization have basic SEO knowledge, the likelihood of success is exponentially greater. The SEO team or consultant can focus most of his energy on new initiatives and continuing to develop the process while the rest of your team makes good business decisions and taps the SEO specialist for on-demand answers to questions.

N O T E

A special thanks to Jessica Bowman (<http://www.seminhouse.com/>) for her contributions to this chapter.

Solutions for Small Organizations

Some organizations are not equipped—either structurally or financially—to have an entire SEO team. In fact, only one person may be knowledgeable about SEO, and that person may be only a part-time employee. Or there may not be anyone within the organization with the time or skills necessary to optimize the site, so outsourcing may be required.

The amount of time that is needed to perform effective SEO work depends on the size and complexity of the website, as well as the competitiveness of the market space the organization is pursuing. An organization's size and vertical scope add to the overall complexity of the SEO process. Some of the ways that SEO affects complexity include:

Keyword research

More pages mean more keyword research. Solid keyword research is needed to help drive the site architecture and content plans.

Titles and headings

As two of the most powerful on-page SEO ranking factors, titles and headings need to be chosen for every page. For very large sites, you may have to design an algorithmic method to choose these for you.

Site architecture

Websites should have pages that address the majority of ways in which potential users think about their products and services; and as we covered in [Chapter 6](#), a larger site is likely to have a more complex hierarchy.

Meta descriptions

Meta descriptions are important because search engines often use an excerpt from your meta description in the SERPs, which influences your site's click-through rate.

Link-building efforts

As sites scale, the complexity and need for links grow. You need to research the competition levels of keywords and make plans for link development so that execution neither grossly exceeds nor falls short of the necessary effort.

Web-based partnerships

Websites of all sizes engage in partnerships and relationships with other entities (charities, businesses, consultants, clients, etc.). SEO professionals know that all of these contain opportunities for acquiring links and link relationships, and that when they are properly leveraged they can result in massive value-adds for the organization. For larger organizations, these partnerships can be more complicated in nature.

Development platforms and CMSs

The development platforms and CMSs used on larger sites can often create a number of limitations regarding SEO implementation, and frequently require extensive, costly, and time-consuming workarounds before optimization efforts can be implemented.

In-House SEO Specialist

Building SEO knowledge in-house can be challenging in a small organization where most of the people involved are already handling multiple functions. It is often good to have an SEO consultant on call to answer questions and validate solutions.

Before assigning SEO to existing talent without the budget of an agency, consultant, or contractor, spend \$300 to \$1,000 to engage an SEO professional to evaluate the aptitude of the person(s) you have in mind for the role. Countless companies have underestimated the specialized skills required for successful SEO. It's a role unlike any your organization has seen—this person needs the technical skills, marketing panache, and political savvy to work his way into website development processes, and the innovative, outside-the-box thinking that will generate creative solutions to search engine crawler needs. Choosing the wrong person can be a costly mistake in terms of time and missed opportunities.

As with larger organizations, it is still important to develop a basic level of SEO knowledge throughout the organization. SEO still touches management, marketing, and development, and it is important that all of these departments have a basic understanding of the issues.

We will outline some of the ways to effectively and rapidly learn about SEO and build up in-house knowledge in “[Working with Limited Resources/Budget](#)”.

Outsourced Agency/Consultant/Contractor

Sometimes the ideal solution is to outsource SEO, and one of the biggest advantages of this for a small organization is that it allows the company to access an expert without having to have an internal resource. You can bring in a strong SEO professional almost immediately, whereas creating one in-house would take at least six months, and more likely a year or longer.

Hiring outside help can work effectively, but be sure to follow the guidelines we will discuss in “[The Case for Working with an Outside Expert](#)” on page 519 and “[Selecting an SEO Firm/Consultant](#)” on page 521 when selecting an SEO firm.

Working with Limited Resources/Budget

Doing SEO on your own and learning SEO can be a challenging task for two major reasons:

- The demanding, ever-changing landscape of search algorithm behavior is often unpredictable and nonintuitive.
- There are literally thousands of details and tactics to learn, some of which may have little impact on their own, but when used in various combinations with other components can hold powerful influence over ranking. Herein lies the “art” aspect of mastering SEO.

Fortunately, many SEO training materials are available via paid subscription at [SEOMoz](#), [SEO Book](#), [Market Motive](#), and the [SEMPO Institute](#), among others. If you don’t have the budget for a subscription, you can try public blogs and resources such as the SEOMoz Blog & Guides and [SearchEngineLand.com](#). Also consider the many resources for learning that we outlined in [Chapter 11](#).

Basic Low-Budget SEO Ideas

You can do numerous things, at a fairly low cost, to improve your site’s overall optimization, including the following:

Use the free search engine tools

Use the free set of tools provided by the three major search engines. Create an account in Google Webmaster Central, Bing Webmaster Center, and Yahoo! Site Explorer, and verify yourself as a site owner on all three. This will provide you with access to diagnostic tools, such as *robots.txt* validators, and reports on backlinks, spidering activity, server errors, top search queries, anchor text, and more.

Check out your competitors

Assess your site and those of your competitors for SEO success factors such as keyword URLs, keyword-rich title tags, H1 tags, keyword prominence, and so on. To survey yours and your competitors' title tags across a large number of pages, use the search engines' site operators and set (in the preferences) the number of results returned per page to 100.

Optimize your title tags

You want each title tag across your site to be unique and focused on a relevant keyword theme. Make each title tag count, because it's given the most weight by the search engines of all the elements on the page—and it heavily influences the searcher's click decision from the search results.

Optimize other critical elements

Analyze the text, HTML, inbound links, internal links, anchor text, and so on to determine your ideal configuration for success (include a dose of your own critical thinking).

Measure, test, measure, and refine

Test your assumptions and the assertions of others—particularly SEO bloggers (not every piece of advice you find will be accurate or applicable). Measure Key Performance Indicators (KPIs) and treat SEO like a series of experiments. Make iterative improvements to your URLs, title tags, H1 tags, internal linking structure, anchor text, page copy, link acquisition efforts, and so on.

What sorts of KPIs should you measure and aim to improve? At a minimum, consider checking rankings, traffic, and conversion metrics. However, you can also check other metrics, such as the number of different search terms used to find your site, the number of different landing pages where search visitors arrive, the growth of inbound links and/or the addition of any notable inbound links, and so forth.

Make nice with IT—or bypass them

If you are in a larger organization, you must persuade your IT team to have a stake in your site's SEO success because you rely on them to implement many of the optimizations (such as URL rewrites, 301 redirects, breadcrumb navigation, tag clouds, PageRank sculpting, etc.). IT can be your biggest ally or the thorn in your side. They need to weave SEO into every facet of what they do, and that happens only when they want to work with SEO. Start by heaping on the praise and appreciation—maybe buying them pizza!

If conflicting priorities and insufficient internal resources stymie you, you will have to institute workarounds that minimize IT involvement. Have no fear: you can still accomplish what you need to with your site, with a bit of out-of-the-box thinking and some technology.

For example, you could use proxy optimization technology (such as Netconcepts' GravityStream) that optimizes the URLs, titles, page content, navigation, and so on via a proxy server, obviating the need for major invasive surgery on the underlying e-commerce platform. Nonetheless, start with the pizzas and go from there.

The link bait option

Get one *great* idea for a tool, service, resource, or page, and bounce it off some folks in the forums (or privately through email to an SEO expert whom you trust). Hire a developer who can help you build it; use Craigslist (<http://www.craigslist.org>) or other online classifieds and post with a budget you can afford.

Leverage low-cost tools

Consider using one of the following tools:

Adobe's Dreamweaver (<http://www.adobe.com/products/dreamweaver/>)

This is a de facto tool for any website/application developer. Most of the time it is used to build web pages, but it also has a very good range of reporting tools. With a few mouse clicks it is possible to:

- Analyze website accessibility (WAI and other standards)
- Analyze website coding practices
- Find empty or missing title tags
- Find empty or missing image alt attributes
- Analyze link structure (orphaned files, external links, and broken links)
- Store everything in nice reports anyone can view in HTML format

To make things even nicer, you can fix most of the noticed errors/warnings automatically, either directly or indirectly. And even better, you can extend the preceding list of available default tools with scripting and extensions.

Xenu's Link Sleuth (<http://home.snafu.de/tilman/xenulink.html>)

This is a simple link-based crawler. Web developers use Xenu to check for broken links on a regular basis, but for SEO purposes the best value comes in the form of simple internal link analysis. By ordering the Xenu Sitemap based on "links in" and "page level," it is easy to detect possible internal linking abnormalities that interrupt PageRank flow or decrease anchor text value; and of course, you can save all this information as a report. Xenu gathers loads of information, and it is a very useful tool, even for in-depth SEO purposes.

Microsoft Word (<http://office.microsoft.com/en-us/FX102855291033.aspx>)

Although it may seem to be an unconventional tool for a web developer/SEO practitioner, it is undeniably one of the best copywriting and publishing tools that all users are familiar with. Microsoft Word has several built-in features that help to produce high-quality content, analyze existing content, fix and locate essential grammar errata, and above all, easily automate and synchronize all features and changes with other users and publishing tools. For more tech-savvy folks, there is always the scripting option for fine-tuning.

As with most SEO tools, the beauty is in the eye of the beholder. If you use the preceding tools properly, they can be very helpful; for the untrained eye or for the wrong kind of task, they can cause pain and misery. Making proper on-page optimization decisions

usually takes tens of hours. Even for a relatively small site, it is possible to cut that down to fewer than two hours with the tools and methods we just described. Of course, there is a difference in the quality of work and the documentation you'll get with these tools compared to what you'd get with more conventionally priced SEO services.

Limited cash options

If you're on a tight budget, you can look into purchasing Steve Krug's *Don't Make Me Think* (\$23.10) or a Yahoo! Directory listing (\$299). You can also check with SEO firms to see which ones are offering special introductory packages. Many firms offer special pricing just to get started with a client.

These are just examples of things you can do, and countless other options are also available. Whatever you do, don't short-circuit the process of developing in-house expertise, even if you hire outside help. That in-house knowledge will be invaluable over time, and if you have hired a quality outsourcing partner they will be delighted as it will make their job easier. They may even help train you or your team.

Solutions for Large Organizations

The challenges of performing SEO for a large organization are a bit different from those for a small organization. This is true regardless of whether you are working from an in-house position or are an outsourced SEO consultant. Some of the challenges can stem from the size of the site, which can range from 10,000 to 10 million pages.

Large organizations are usually complex entities, and many of these organizations may make these decisions by committee, or review all decisions with a committee before finalizing them.

Many important projects (SEO or otherwise) can be delayed or even canceled in a large organization due to a lack of understanding by a single key player. In the world of SEO, efforts are often delayed by someone in IT/development, marketing, sales, or management—and it can be challenging to get a meeting with the people you need to persuade, let alone accomplish the task of persuading them.

Patience and persistence are essential, and adhering to the following large-organization SEO guidelines is recommended:

- Get buy-in, if you can, from the head of IT, the head of marketing, the head of sales, and senior management. Those people will likely control your fate. If you can get two or three department heads and senior management on board, you should be in good shape.
- Always make sure you talk about opportunity cost. Make sure all people involved understand they're leaving X visitors on the table every day, and that at the current conversion rate that means N potential leads and Y potential dollars. Include the *opportunity gap*—where the client is versus where they could be—in every report.

- Insist on a sound web analytics plan. Successful SEO projects depend highly on quality analytics information. If your company is stuck in this situation and can't get out, yet you have support for implementing SEO, look at other KPIs, such as rankings combined with search volume and estimated clicks-per-position, in the search results.
- Provide detailed reports, even when you're not asked to. As the disciplined people around you are likely to be most unfamiliar with the basics of SEO, SEO practitioners need to overcommunicate.
- Educate, educate, educate. Always explain why you are making a given recommendation. Just make sure you are speaking in the language of the recipient (i.e., talk technically to the developer and in business terms to the marketer). This is where many companies fizzle out. When you conduct training properly, you can create an almost overwhelming amount of buy-in and interest in SEO. Companies often go wrong when they think one training session is enough. Unfortunately, interest will wane, and if you don't offer continuing education you can be back to square one in 6 to 12 months.
- Be confident. Cover your bases before each meeting. Don't waffle. The corporate boardroom can smell fear. They'll question your judgment anyway, so make sure you don't give them any additional ammunition. It's important to speak confidently on what you know, but it's also important not to speak about what you do not know. There is nothing wrong with saying, "That's a good question. I don't have an answer for you right now. Let me do some research and get back to you." Most development teams will respect that answer if you can do your research and get back to them quickly. Be mindful of the Ugly Baby Conundrum: the website is the baby of programmers, designers, and business sponsors; when you talk badly about the site, it's like you're saying, "Your baby is ugly." When you look at a site for SEO, you will probably find many issues. Be good about noting what they did correctly for SEO, and be unbiased when presenting the issues you've found.
- Don't make super-technical presentations to the executive team. Although they might understand what you're saying, their brains are tied up with 99 other things. Just get to the point: what's gone well, what hasn't, and what you need from the team to fix it.

Contracting for Specialist Knowledge and Experience

Even if you have a solid in-house SEO team, sometimes it makes sense to get help from the outside. Why would you want to bring in external expertise when you have a strong team already?

The answer lies in the complexity of SEO itself. SEO has dozens of subdisciplines, including video optimization, local search, image optimization, competing for search traffic in other countries, widget promotion, link development, and strategies for social media properties.

Another reason to bring in an SEO consultant is to conduct an audit on your site. This will identify new opportunities, identify the things you missed, and confirm the things you did right. Here are the ideal times to engage a consultant to perform an SEO audit:

- When you need to understand the amount of work involved in optimizing your site to get approval
- After you have done what you know about SEO and need advice on what to do next
- When others in your organization need justification of your SEO recommendations
- Before a site redesign so that you can learn what needs to change and what you can do to make the next website more search-engine-friendly

Applying SEO Recommendations Intelligently

One of the biggest challenges in SEO is the gap between strategy and implementation. The SEO artist knows how to explain his recommendations in simple terms and in the language of his audience. However, simple expression of a concept does not always mean seamless implementation.

When you look at a site from an SEO perspective, you must always keep the bigger picture in mind: to do what is best for the business by balancing both short- and long-term goals while managing available resources effectively. Changes that minimize implementation cost and hassle but have a big impact have the best ROI. Before you recommend throwing out that old CMS and reinvesting millions of dollars in a brand-new website, ensure that the ROI to the organization will justify the means, and back it up with data.

This does not mean you should never recommend drastic changes, but rather that you should do so at the right time and place. Be mindful of the potential ROI for each SEO recommendation. Sometimes this is difficult to quantify, and you may end up with a table similar to [Table 12-1](#).

TABLE 12-1. Rating potential SEO projects

Project	Value	Effort
SEO Project 1	High	High
SEO Project 2	Low	Medium
SEO Project 3	Medium	Low
SEO Project 4	High	Low

This type of analysis is incredibly valuable in keeping an organization moving forward with an SEO program and increasing search traffic at the most rapid rate possible.

Hiring SEO Talent

If you have decided that you want to add to your in-house SEO team (or start one), you need to think seriously about the type of person you want to hire. Do you want someone with prior experience? Or do you want the person to learn from other team members along the way?

Hiring top people can be prohibitively expensive, because there are a lot of income opportunities for them to pursue. It is hard to match the earning potential of a top SEO practitioner because, at the very least, he can make more from building his own sites and operating them in an affiliate model. That does not mean you cannot hire real talent; it just means you will pay a premium.

How to Select the Right SEO Practitioner

Although experience costs more, under normal circumstances it will bring faster results. On the other hand, a less experienced person is a lot like a rookie in baseball: if you hire a future star, you can potentially accomplish a great deal at a much lower cost.

When sifting through résumés, examine candidates' specific skill sets. If you are looking for someone to develop social media campaigns, you should probably be a bit more focused on that than on whether the candidate is an expert at keyword research. No matter what the person's skill level, he should "get" the Internet. Does he blog? Tweet? Participate actively in social media sites such as StumbleUpon? If he doesn't, chances are he is underexperienced for the position you are looking to fill.

When evaluating a person's skills, don't forget to assess his political finesse. What makes a successful SEO practitioner at an agency or in a freelance position is different from what makes a successful candidate in-house. Often, companies focus strictly on results and types of projects; however, if your company is plagued with red tape and political minefields, you need someone who can sell SEO to the entire organization and integrate it into the organization in the right places so that it is implemented successfully. There is an 80/20 rule for in-house SEO: 80% of the time you spend selling and 20% of the time you spend doing SEO. These skills are very different from what is required as a consultant, where you spend 80% of your time doing SEO.

Pitching the Person

Finding good SEO talent can be difficult and challenging as top SEO experts are in high demand and likely have more than one job option to pursue.

Many companies overlook this. The "jobs offered" sections of forums are often full of offers, but few have made any real effort to market the opening. It really is a basic marketing problem that many of those vacancies show. Many don't even put a location in the title. Far worse is that almost all of the posts are company-centric: all about what the company wants from applicants, and no thought for what the prospective applicant will want from a company.

Fail to pitch the prospect and you will get to choose from a diminishing pool of folks who can't make it alone, aren't able or willing to make a good living from affiliate programs, and don't prefer to be with a more forward-thinking and proactive company that looks to develop its staff, rather than merely recruit them.

Sample Job Opening

The following job opening appeared in the job postings at Search Engine Watch (<http://forums.searchenginewatch.com/archive/index.php/t-4088.html>):

I'm totally serious about this unique opportunity for a person of the right calibre to come and work with Propellernet UK, a Brighton-based Search Engine Marketing company, as a trainee Consultant.

This is not your average trainee position, and a high level of skills are already expected. We do not have the time to spend years on this training, and we are looking to take on an experienced SEO, who already understands the special aspects of Marketing as it applies to the web. Then I will personally train and coach you to take that further still.

The ideal candidate will have a background in sales or marketing, as well as a broad understanding of SEO.

(Clue: If you don't know who I am, your knowledge of SEO is not going to be broad enough.)

You must be a great communicator, as an essential part of this role is to bridge the communication gap that often exists between the Marketing and Technical Departments of Client Companies.

You need to be able to talk to Marketing Directors on their level, and then turn around and talk with Heads of IT Departments on their level without a pause.

You need to have equal passions for both Teaching and for Learning, as you will be expected to do both things each week as a matter of course.

You need high initiative, and the ability to think on your feet. You need to be proactive yet still have fast reactions too.

You must be in possession of a full, clean, UK Driving License, as considerable travel may be involved in the consulting work.

What you get in return:

A salary based upon ability and experience that will be reviewed regularly as your skills are proven. The salary will also include performance-related elements, and on-target earnings could easily surpass £70,000 as soon as your skills are up to the demand.

You get to work right on the cutting edge of SEO consultancy, where the work is never dull nor repetitive and we seek unique solutions for every client.

You get to be trained to be a world-class SEO and Web Marketing Consultant, able to stand alongside the very best that our industry has to offer.

I am looking to train an equal, not a junior, so you'll be learning everything that I can impart to you just as quickly as you can absorb it.

This is a completely sincere offer. Your time will not be wasted - I don't expect mine to be either.

CVs should be sent to Propellernet Ltd. - you can find the address for yourself if you're a qualified candidate.

Sincerely,

Ammon Johns (aka Black_Knight)

Marketing Director, Propellernet Ltd.

Clearly, this type of job posting is not what every company would want to put out there, but that is part of what makes it effective. There is a certain amount of attitude to it, and that speaks volumes to a potential applicant: either this is what you are looking for, or it isn't.

In addition, the posting makes it clear that Ammon Johns is looking for someone who can be a true peer. This and many other points in the job outline make it clear that he is looking for a senior person. One thing that might make this a bit more effective is the addition of some examples of search problems that a qualified applicant should be able to solve prior to applying (the "find the address for Propellernet Ltd." requirement at the bottom is not a challenging problem).

The job requirements also spell out many of the personal and professional attributes an applicant would need to have to get the job. The compensation is well detailed too, so the applicant knows what is in it for him.

The Case for Working with an Outside Expert

You might want to work with an outside expert if any of the following scenarios apply to you:

- You don't have any inside SEO resources, are not in a position to hire them, and need someone to work directly with your development and marketing teams to get it going.
- You don't have any inside SEO resources and you want to build a team, but you need to jumpstart the process in the meantime. You may also want to use your outsourced SEO practitioner to help train your team.
- You have an inside team but it is missing a key piece of expertise, or you just want to sanity-check the conclusions your team is drawing or see whether they are missing anything. In addition, since SEO at its best involves lots of brainstorming, an injection of fresh ideas can be just what the doctor ordered.

Regardless of the scenario, an outsourced SEO professional can bring two assets to the table that will be difficult for your internal team to provide:

The power of external authority

If you hire an industry-acknowledged expert, he can often speak with a voice of authority that is not readily available to the inside team. This is not necessarily because the inside team is less senior; it's just because they are, well, inside. Familiarity may not be breeding contempt in this case, but it can still make someone seem less of an expert, perhaps because that person's occasional mistakes have been seen.

In these scenarios, an outside expert can come in and not be subject to that bias. This could be what you need to convince someone in senior management, marketing, or IT about the right way to view a key issue and get things moving. From management's point of view, the opinion of a recognized authority can increase their comfort that the right path is being pursued. It is very common for in-house SEO professionals to engage an outside expert to help their cause. Engagements range from a presentation or training to site audits and quick consulting calls with IT.

The value of experience on multiple sites/campaigns

Outside experts will have the benefit of having worked on many different sites and campaigns. Each campaign is like a lab experiment in which the SEO expert can benefit from many different types of tests to learn what works well. In addition, if your outside expert is active in the SEO community, he may also benefit from discussing different experiences with other SEO experts.

How to Best Leverage Outside Help

As we discussed previously, you might bring in an outside SEO expert in three scenarios. In all three scenarios, you must get someone who is good with people, can negotiate, and can teach. Bringing in someone you can trust is also a must.

To leverage these skills you need to put the person in a situation where he can succeed. An SEO expert will tell you what he will need from you to be successful. Take this input seriously. If you can't provide the things he is asking for, you have no chance for success. It is vital that management paves the way for successful SEO plan implementation. Although many managers may believe they are giving SEO projects what they need, this is not always the case, and as a result, SEO suffers.

For example, the SEO expert may need time from development to implement suggestions he makes. If the development time is not available, his suggestions are useless to you.

On the content side of things, the SEO practitioner may define a content plan to support his link-building campaigns. Before you hire him, make sure you have a rough idea of the scope of the content development he thinks might be needed to support his link-building ideas.

Take this input seriously, because this is the SEO expert telling you what he thinks he needs to be successful. If he doesn't get that support and ends up thinking he can't win, what are the chances he will succeed? The thing to remember about SEO is that implementing 20% of the recommendations rarely equates to as much as 20% of the potential lift. Unfortunately, most companies implement only a fraction of what is needed and wonder why their traffic isn't increasing.

Ultimately, the cost of hiring an outside SEO expert goes beyond the direct cost for the expert's services. The SEO expert will need a certain amount of support to have a chance for success. Before you decide to hire an SEO professional, ensure that you have the internal support (and are willing to provide it) to make the recommended changes. You don't want to be one of those companies that pays for SEO advice that sits on the shelf and collects dust.

Selecting an SEO Firm/Consultant

Once you have decided to hire an outside firm, you need to go through some sort of selection process. This can be daunting if you are not already familiar with the space and if you don't already have some understanding of SEO on your own. There are many solid and reputable companies, but unfortunately there are a large number of bad ones as well.

In addition, there are many different types of projects you may want them to work on. For example, you may have a local SEO-focused project in mind. This obviously becomes one of your major requirements, and you need to find a team with the right type of experience. Similarly, you may have specific contract requirements that will impact which firm is right for you.

Getting the Process Started

Just like a bad hire can be a very expensive proposition, so can engaging the wrong consultant or firm through a haphazard process or by simply relying on a gut feel. Here is an outline of steps for a more methodical, disciplined search for outsourced SEO expertise:

Start with your goals

Connect with your team and make sure you have a good idea of what you want to accomplish by working with an SEO expert. Sometimes it's as simple as an audit to make sure you're following best practices; other times it's training for your marketers and content builders to get them up to speed on how to actively promote the site. If you have a solid list of things you want completed at the end of an engagement, you'll be better able to judge the eventual results.

If you don't really know what you want, that's OK, but it does make the SEO practitioner's task a little less structured, and it means you need an exceptional level of trust in him. Sometimes an SEO expert can help you define your goals—to get the best of both worlds,

you can ask him what he thinks your goals should be and combine those with your own list.

Connect with your social network

The people who can best assist you in finding a good match are those who know you well. Talk to friends, fellow business owners, SEO bloggers, and personalities that you know and trust.

When you talk to your network, make sure you share a little bit about the project and the goals. The more information you can share, the better folks are able to assist. At a minimum, if you have geographic considerations or want a company versus a freelancer, this is good to include in the request.

Get advice from SEO-savvy people you trust

Look seriously into familiarizing yourself with forums, blogs, and sites related to SEO and the people behind them whom you respect. It's easy to establish a relationship by asking for advice on which SEO company or consultant to hire. As a rule, SEO experts are incredibly friendly about referring business to good companies and good people, and if your network has already given you a few names, running these by the professionals can get you valuable insight on potential winners versus lemons.

Make sure the person is available

This is a fairly obvious step, but many people do not put this at the beginning of the communication process. Ask the person whether he is available at the start, and then continue the conversation if applicable (there is no need to waste energy on folks who can't help, though you should always ask for a referral). Your time (and theirs) is precious, so don't waste it.

Preparing a Request for Proposal (RFP)

A well-written RFP can significantly improve the overall process. If your RFP is not written well, it hinders the SEO firm's ability to understand and define your needs and to scope and price your project. This in turn leads to a disconnect in expectations for both parties. A lousy RFP can discourage a busy SEO firm from even responding—which is a very unfortunate outcome, since it takes the best firms out of the running.

In addition, some top SEO firms are reluctant to participate in an RFP process because RFP responses often require that the SEO firm provide a substantial amount of free consulting just to bid on the job. So, it can sound like a good idea to ask for a lot of great information in the RFP, but this can become an undue burden on the RFP respondents.

Many companies intuitively “know” what they want but are challenged structurally to “ask” for it in a way that is clear, succinct, informative, and constructive. If written properly, an RFP will facilitate the sales process and ensure that everyone involved on both sides understands the purpose, requirements, scope, and structure of the intended engagement. By following a

few key steps in the beginning of the RFP process, you will be able to rest easy, knowing that you are going to get what your company wants in the way that is best for you.

Step 1: Nominate a “point person” for the engagement

One of the most challenging concepts for any large company with multiple working parts is to determine the end goal of the engagement. Often, marketing departments may voice different wants and needs than IT departments; even when they are asking for the same thing semantically, they are not using the right language to communicate what they are looking for.

Every successful project needs a champion who is invested in the project’s success and can pull together the disparate groups that have a stake in the outcome or a role to play. Without that person at the helm, the project will struggle. By nominating that champion as the SEO firm’s “point person” even before you send out your RFP, you will ensure a steady flow of information throughout the process so that internal and external expectations are met.

Step 2: Define “needs” and “wants” using a decision matrix

RFP recipients will understand that you aren’t the SEO expert, and therefore you aren’t going to be able to adequately define the scope of your desired SEO engagement. As the saying goes, “You don’t know what you don’t know.” Just try to be reasonable when formulating your needs and wants and recognize that the RFP recipient may know what you need more than you do.

In addition to disclosing to the SEO firm the basics such as your objectives, your site’s *conversion event* (e.g., online purchase, newsletter subscription, white paper download, etc.), target audience, constraints, and so forth, in all fairness you will also need to clearly spell out the criteria for which you will evaluate the SEO firm’s suitability. This means you will need to define these criteria in advance.

SEO firms know the selection process is usually governed by gut feel, so any RFP recipient who does not have a preexisting relationship with you is automatically inclined not to respond because he knows the deck is stacked against him. This is further exacerbated when you send the RFP to multiple SEO firms. You can allay this concern by candidly sharing with the firms your biases and the clearly defined criteria by which you will be evaluating them.

You should map this out into in a *decision matrix*, which is simply a chart listing the attributes you are looking for in the SEO firm, a weighting factor for each attribute, a score from 0 to 10 for each, and the weighted score (e.g., the score multiplied by the weighting factor). The weighted scores are added to arrive at a total score.

Judging criteria that are both quantitative and qualitative brings objectivity into a subjective process, which will aid you in managing expectations internally. And by sharing your list of criteria and the weighting factors with the RFP recipients, you increase the likelihood—as well as caliber—of responses.

Step 3: Define your success metrics

Now that you've identified what you are looking for in an SEO firm, it's important to let the firms know how you will measure the success of the engagement. Some firms can report on SEO health metrics well beyond just rankings (e.g., page yield, keyword yield), thus facilitating troubleshooting and reviews of program performance. Others will rely heavily on your own analytics package to track the program's success.

For example, you might pose this question internally: "What are your KPIs?" SEO firms often use KPIs to prove the value of the services they provide; online retailers often get caught up with conversion, a metric that isn't really under the SEO firm's control. If you aren't sure which KPIs your company would like to use, think about baking that request into the RFP, stating something such as, "Do you offer monthly program performance reviews indicating the program's growth? If so, please elaborate on the deliverables of such reviews."

Step 4: Prepare to disclose all known influencing factors

An SEO firm can easily examine your existing website—the on-page factors (title tags, navigation, HTML)—to gauge project scope. But other factors will influence your rankings during the engagement, some of which won't be immediately obvious or known to the RFP recipient without prior disclosure. Does your company have other domain names, subdomains, or microsites? Is a massive redesign of your website in the works? Do you employ a third party for your site's internal search, and if so, who is it? By disclosing as much information as possible, you spare the SEO firm the time and expense of discovering these things on its own.

Step 5: Provide an estimated timeline and budget for project completion

One of the biggest deterrents in any RFP is the confusion over when a project should be completed and how much it might cost. A company hiring an SEO firm may not know how much time it takes to complete an SEO audit, but it may have pressing internal matters that require a specific deadline.

Retail sites may want to schedule new launches around a particular theme or season; for example, if you want your site launched in time for back-to-school, be sure to write that in your proposal. Not only will an expected deadline save your company time in looking for an appropriate firm, but it also serves as a professional courtesy to the SEO firms you are querying.

Budgets often coincide with a project deadline, depending on how your company conducts its business. Consider assessing a budget based on a range of services; that is, instead of saying the project absolutely has to cost \$X, say you are willing to spend within a range of \$X to \$Y. Budgets may or may not be included as part of the RFP, and there are benefits (and drawbacks) to both approaches. Even if you do not provide an SEO firm with your budget in the RFP, we recommend determining a budget beforehand because the money you are willing to spend will help you determine a target ROI.

A sample RFP document outline

Now that you've gathered the preceding information for your RFP, you're ready to sit down and write it. You can structure an RFP in several ways. Here is a brief outline:

Section 1: Summary and overview

This section is where you will introduce the challenge you are having and provide the SEO firm with an overview of the rest of the RFP. We also recommend outlining how you would like SEO firms to respond to your RFP. Think of this section as an executive summary, where you will provide the highlights of the RFP without the technical details.

Section 2: Technical summary

Often, SEO firms need gritty details to determine how they can best help you. In the technical summary, this is where you will provide key pieces of information relevant to your project, such as technical requirements, a description of technical issues with the project, your site's current platform, and so on. The technical summary might originate from your IT department, as this section is often for an SEO firm's programmers and delivery team.

Section 3: Administration and management

By describing who will be involved in the project on your end and what the timeline is for completion, you are finalizing the framework of the project.

Section 4: Project expectations and delivery

If you've done your homework, this should be the easiest section to write. From outlining your evaluation criteria to outlining what monthly deliverables or training you would like to receive to assessing your KPIs, this section helps an SEO firm determine its cost to complete your project, as well as its suitability.

RFPs can be pretty daunting to write, but if you think about it, there are several benefits to creating one. Not only will your company identify potential barriers to project success, but you'll also facilitate great internal communication, develop better budgets, and identify what types of SEO firms you'll want to work with.

When you write an RFP, keep in mind that the purpose of an RFP is to hire an expert SEO firm that will propose its recommended actions for your company's website(s) to achieve the greatest chance for success. Sometimes an SEO firm's feedback highlights other potential issues that you may not have considered, which may change the scope of your project entirely.

If you don't know what the best solution is for your web properties and have only a vague idea, you can submit a Request for Information (RFI), which is typically not contractually binding but allows you to ask questions about an SEO firm's products and services.

Communicating with Candidate SEO Firms

You should also plan to talk to candidate SEO firms. This is your opportunity to dig a little bit deeper and move beyond the RFP. If you are a smaller company and have bypassed doing an

RFP, you will still need to be in direct contact with your prospective SEO firm. Here are some things for you to consider during this stage of the process:

Ask for a list of past success stories (not just clients)

Let the company tell you where it has helped someone be successful. You can ask for a client list, but be aware that a failed client engagement does not necessarily mean the SEO firm failed. Many times clients don't implement the SEO practitioner's recommendations, or they try to, but do it badly.

You can use this to your benefit. Ask them for a case study of a client failure, and then ask them why they think it failed. This can provide some great insight into what they require from their clients, because the failure they disclose to you will inevitably be due to the client not providing them what they needed.

Talk on the phone or (if possible) get together

Email is a great initial communication medium, but a phone call or in-person meeting gives you a real sense of the team you'll work with. Make sure you're not talking to a salesperson (if you go with a big company), but rather to the person who will be interacting with you throughout the contract execution process. There's nothing worse than getting sold by a charismatic, knowledgeable SEO leader and then being passed on to a junior team for management.

In the phone call/meeting, be sure to establish a rapport. If you can't feel comfortable on a personal level, don't dismiss it—brilliant SEOs are great, but if you can't work well together, the project is unlikely to succeed. Everybody works better with people they like and people they get along with. There's the additional pressure of not letting down someone you've developed a relationship with, and it extends subconsciously and consciously into every part of the work you do together.

Present a few initial issues over email

Some SEO consultants won't engage in any work before a contract is signed, but you should at least get them to talk about how a problem can be approached, whether it's an issue, and what strategies they might recommend to fix it (even from a broad level). If you are an SEO practitioner, put a few examples of things on the table for the potential client. It builds trust in the SEO practitioner's knowledge and gives you insight into how he solves problems and what the work relationship and advice will be like down the road.

References

Asking for references is a good idea. Make sure you qualify the reference as part of the process. You want to protect yourself from people who might get some friend of theirs, instead of a real client, to act as the reference. It is helpful if the reference the candidate provides works within a well-known business. Make sure you include an understanding of the relationship between the SEO firm and the client provided as a reference.

Once you get the reference on the phone, you should try to learn what the SEO consultant was asked to do, and what he did. Get specific details on this. If you want someone to perform link building for you but he did not do link building for the client, the value of

that reference is decreased. Then see whether the reference will provide you some input on the benefits they received from the work of the SEO consultant (e.g., increases in traffic, conversions, or other metrics of value).

Also, think about conducting your own sleuthing. The [Wayback Machine](#) can show you prior versions of the reference's site, including client lists. Call ex-clients the candidate firm didn't give as references. In addition, ask about good and bad experiences with the company in forums that cover SEO, such as Google Webmaster Groups.

Making the Decision

Regardless of whether you have done an RFP or used a more informal process, make sure you understand the process you will use to make a decision. Here are some tips on how to do that:

Get a written response from two or three vendors

Don't get an RFP response or quote from just one vendor, even if you're sure you're going to use them. Multiple quotes aren't just a best practice, they're a good way to learn about pricing and scope discrepancies. The quote from the vendor you don't like might have some pieces that your preferred vendor overlooked. Pricing is harder because it fluctuates so wildly in the search marketing world.

However, don't choose an SEO expert on price alone (or even make price the biggest part of your decision) unless your budget is a real problem. SEO is usually an incredibly high-ROI activity. Companies that spend larger amounts on SEO services tend to make it back in a matter of weeks or months from traffic and conversion increases (remember that more targeted traffic means higher conversion rates too).

Thus, going with an SEO expert who costs a lot less might seem like a good idea, but if you don't work as well together and you think he might not do as good a job, you're hurting yourself in the long term. Of course, don't get fleeced by an exorbitantly overpriced firm (watch out; some of the biggest SEO companies have some of the most ruthless pricing models because they know that Fortune 1000s will work only with consultants who've done work for other Fortune 1000s), but be aware that a few thousand dollars is not the best reason to choose a different provider.

Have smart, sensible people review the contract

Lawyers are great, but sometimes legal folks get overly concerned with liability and risk management details and overlook big-picture business items. Make sure your savvy business/ROI-focused personnel get a good look too. If there are legal issues that cause a rift, it can even be wise to get some C-level executives involved. Sometimes the folks from legal won't budge on an issue that's going to be a deal breaker, but if the CEO says do it anyway, you can get around the more problematic demands. Also be aware that different things are important to different firms. For example, some SEO companies will walk away from a deal if the client insists on a no-publicity clause.

Go with your gut

When you finally make your choice, go with the team you feel will bring results. Making a matrix of price versus service versus reputation versus estimated productivity is fine if you're into that, but consider going with your initial gut feeling. If you don't feel like there is a correct company at the time, go back to the table, get more information, and even try to get another bid or two. When you do find the best choice, things just "feel" right. How do you know?

- The way they talk about search engines makes sense and the advice they've given lines up with the best practices you've seen expressed on SEO sites you trust.
- Their level of familiarity with recent trends at the engines and the links they send or news they cite is timely, relevant, and logical.
- They use the engines like pros, rapidly combing through site queries, link information, and analytics data to unveil the underlying problems that are hurting rankings or causing the competition to succeed.
- Their interactions with your team "fit." The team members like the consultants and are eager to implement their advice.

Remember, an SEO business relationship is a relationship nonetheless, and if the two parties involved simply speak different languages professionally and/or culturally within their organizations, no amount of SEO expertise—and no amount of SEO budget—will make the relationship work. It is much smarter to work with people you like than to pay begrudgingly for expertise you need that comes in a package you don't care for.

Mixing Outsourced SEO with In-House SEO Teams

When you are combining the resources of an external team with those of your in-house SEO team, there is a need to determine who does what, who decides what is to be done, and so on. It is best to have this all clearly defined up front, or else there will be conflicts and confusion, which will slow down the process of increasing your organic traffic.

If we look first at the process of defining who does what, there are a few considerations, such as the following:

- What areas of expertise does your in-house team have? How does the outsourced resource complement or reinforce that?
- Is the outsourced resource bringing a unique new area of expertise to your organization or is it a supplemental resource to an existing area of expertise?
- Is the outsourced resource going to be working on an area where you consider the expertise essential to build in-house?

- Can you leverage the outsourced resource with your internal resources? Perhaps you can use this to keep the outside help working on higher-value tasks that provide you a bigger return on your consulting dollar.

All of these considerations factor into the overall division of labor between your contracted team and your internal team.

You also need to be clear about who is making the decisions and who is communicating those decisions to the outside SEO consultant. He will need guidance and direction on what you are looking to accomplish. Although you hopefully defined this in detail up front, it does change over time, so updating that guidance on a regular basis is important. The outside consultant may not sit in your offices all day long, so more effort is required to communicate changes in plans to them than to your internal team.

In addition, the outsourced SEO consultant will have questions and will make recommendations, and he needs to know who to pass these on to and who will give him any relevant decisions that are made or any related updates. A clear communication channel is critical to success.

Building a Culture of SEO into Your Organization

SEO is a marketing function, and like any other marketing function, its goal is to help your organization grow and prosper. For a business, this means bringing in more revenue. What makes SEO different from any other function in your organization is the depth and breadth of its influence on every other role that influences and/or touches the website.

SEO is also a bit of a strange beast, because much of it is nonintuitive. As an SEO practitioner, you can explain things to people and then they will look at you like you have three heads and say, “Huh?” Building an SEO culture requires that you get the organization as a whole to accept a few really important facts:

- Organic search engine traffic is essential to the growth and success of the organization.
- Search engines impose some specific constraints on how you can structure your website.
- Search engine success requires specialized marketing programs.
- Embrace these limitations and requirements, and you can gain a strategic advantage over your competition.

These concepts are pretty high-level and very simple by design. If you can get all parts of the organization buying into them, though, it will have a very powerful impact on your chances for success. Once everyone agrees that SEO is needed, you will have taken the first step in building an SEO culture.

The value of this is that you will have all team members aware, at least at a basic level, of how SEO affects them. Before the senior manager makes a sweeping policy decision about something, she may pause to consider the SEO impact and get the advice she needs on the

topic. Before the developer implements something a particular way, he will likely make sure he is doing the right thing from an SEO point of view.

To build an SEO culture at your organization, you need to institutionalize SEO by integrating SEO into everyday business activities and making it part of everyone's job description.

Successful SEO will not be the result of your in-house SEO practitioner(s); it will come from collaboration with everyone who influences and touches the website.

Conclusion

As we have outlined throughout this chapter, SEO touches IT/development, marketing, sales, and management. Getting every one of those groups to understand the basics of SEO is essential to long-term success. SEO is not the be-all and end-all, but it is a significant component of all these disciplines.

An SEO culture is focused on growing relevant, converting traffic to the website and the business overall. Educating your organization in these simple concepts will have everyone working toward the same objective and will bring a lot of efficiency to your overall operations, which will ultimately help you achieve SEO success.

CHAPTER THIRTEEN

An Evolving Art Form: The Future of SEO

As we have noted throughout this book, SEO is about obtaining new customers.

Therefore, the role of the SEO professional is to bring more targeted traffic to an organization's website. This underlying role will not change as technology evolves, but the mechanics of how it is done will continue to change as technology improves and adapts to users, and vice versa. Driving this is the fact that the complexity of search will continue to grow, as all search engines, not simply Google, seek to locate and index all of the world's digital information. It is reasonable, then, to expect various levels of expansion and evolution in the following areas:

- Indexation of multimedia content, such as video and audio
- Spidering of data behind certain types of forms (something that Google already does in some cases)
- Extraction and analysis of Flash and AJAX-based content
- Local search
- Dominance of mobile search, as the world begins to increase its demand for this capability

Looking a bit more into the future, one can also expect that mobile search will drive an increasing demand for linguistic user interfaces, including voice-recognition-based search. When these things will arrive is not easy to predict, but a lot of people are investing in the development of such technologies.

Business deals also regularly change the landscape. For example, on July 29, 2009, Microsoft and Yahoo! signed a far-reaching deal that will result in Yahoo! retiring the search technology that powers Yahoo! Search and replacing it with Microsoft's Bing technology (<http://www.nytimes.com/2009/07/30/technology/companies/30soft.html>). However, such deals are subject to regulatory approval; this can take some time and a positive outcome is by no means certain. If approval to proceed is granted, the implementation will likely take an extended length of time. It may be two years before the Bing rollout is complete across Yahoo!'s portfolio of web properties.

Many contend that this deal will result in a more substantive competitor for Google. With Microsoft's deep pockets as well as a projected combined market share of 28% by some estimates, Bing will make a formidable competitor. Questions will emerge too. For example, will Microsoft continue to support Yahoo!'s paid inclusion program (SSP), Yahoo! Site Explorer, and the `linkdomain:` operator; or will these invaluable tools fall by the wayside?

These developments and many more will impact the role SEO plays within an organization. This chapter will explore some of the ways in which the world of technology, the nature of search, and the role of the SEO practitioner will evolve.

The Ongoing Evolution of Search

Search has come a long way, and will continue to progress at a faster and faster pace. Keeping up with these changes, the competitive environment, and the impact of new technology provides a challenge and an opportunity.

The Growth of Search Complexity

Search has been evolving rapidly over the past decade. At the [WSDM conference](#) in February 2009, Google Fellow Jeff Dean provided some interesting metrics that tell part of the story:

- Google search volume had grown 1,000 times since 1999.
- Google has more than 1,000 times the machines it had in 1999.
- Latency dropped from less than 1,000 ms to less than 200 ms.
- Index update latency improved by about 10,000 times. Whereas updates took Google months in 1999, in 2009 Google was detecting and indexing changes on web pages in just a few minutes.

These are staggering changes in Google's performance power, but this is just part of the changing search environment. Some of the early commercial search engines, such as Web Crawler, InfoSeek, and Alta Vista, launched in the mid-1990s. At that time, web search engines' relevancy and ranking algorithms were largely based on keyword analysis. This was a simple model to execute and initially provided pretty decent results.

However, there was (and is) too much money in search for such a simple model to stand. Spammers began abusing the weakness of the keyword algorithms by stuffing their pages with keywords, and using tactics to make them invisible to protect the user experience. This led to a situation in which the people who ranked first in search engines were not those who deserved it most, but were in fact those who understood (and could manipulate) the search algorithms the best.

By 1999 Google launched, and the next generation of search was born. Google was the search engine that most effectively implemented the concept of citation analysis (or link analysis) as part of a popular search engine. As we outlined earlier in the book, link analysis counted links to a website as a vote for its value. More votes represent more value, with some votes being worth more than others (pages with greater overall link juice have more juice to vote).

This created a situation that initially made the job of the spammer more difficult, but the spammers began to catch up with this advance by purchasing links. With millions of websites out there, many of them with little or no revenue, it was relatively easy for the spammer to approach a site and offer it a nominal amount of money to get a link. Additionally, spammers could implement bots that surfed the Web, finding guestbooks, blogs, and forums, and leaving behind comments with links in them back to the bot owner's site.

The major search engines responded to this challenge as well. They took two major steps, one of which was to build teams of people who worked on ways to detect spamming and either discount it or punish it. The other was to implement an analysis of the quality of the links that goes deeper than just the notion of PageRank. Factors such as anchor text, relevance, and trust became important as well. These factors also helped the search engines in their war against spam.

But the effort to improve search quality as well as fight spammers continued. Historical search result performance, such as how many clicks a particular listing got and whether the user was apparently satisfied with the result she clicked on, are metrics that many believe have already made their way into search algorithms. In 2008, then-Yahoo! Chief Scientist Jan O. Pederson wrote a [position paper](#) that advocated use of this type of data as follows:

Search engine query logs only reflect a small slice of user behavior—actions taken on the search results page. A more complete picture would include the entire click stream; search result page clicks as well as offsite follow-on actions.

This sort of data is available from a subset of toolbar users—those that opt into having their click stream tracked. Yahoo! has just begun to collect this sort of data, although competing search engines have collected it for some time.

We expect to derive much better indicators of user satisfaction by considering the actions post click. For example, if the user exits the clicked-through page rapidly than one can infer that the information need was not satisfied by that page.

In May 2007, Google made a big splash with the announcement of Universal Search. This was the beginning of the integration of all types of web-based data into a single set of search results, with data from video, images, news, blogs, and shopping search engines all being integrated into a single search experience.

This particular initiative was not directed at spammers as much as some of the other changes, but it was an acknowledgment that there is far more data on the Web today than simple HTML text. The need to index and present that data in consumable format was critical for the search engines to tackle. Google kicked it off with its announcement, but the other search engines quickly followed suit.

Ask also made an interesting contribution with its Ask 3D effort, which it launched in June 2007. This approach provided many different types of results data on search results pages, much like Google's Universal Search, but these elements were not mixed in with the web results, and instead appeared on either the left or right rail of the results page. Google, Yahoo!, and Bing have all borrowed some ideas from Ask 3D.

Search engines also can make use of other data sources, such as registry data to see who owns a particular website. In addition, they have access to analytics data, data from their web search toolbars, and data from free Wi-Fi and Internet access distribution to track actual web usage on various websites. Although no one knows how, or how much, the search engines use this type of data, these are additional information sources at their disposal.

Search engines continue to look for more ways to improve search quality. Google has launched efforts toward personalization, where it can look at a user's search history to get a better idea of what results will best satisfy a particular user. In 2008, Danny Sullivan summarized this entire evolution into four phases (<http://searchengineland.com/danny-sullivan-tackles-search-30-and-40-in-smx-west-keynote-13495>):

- Search 1.0: keywords and text
- Search 2.0: link analysis
- Search 3.0: integration of vertical results
- Search 4.0: personalization

So, what will make up Search 5.0? What is coming next? Many people believe that use of social media data is the next big wave. The "wisdom of the crowds" will become a big factor in ranking. Mike Grehan talks about this in his paper, "New Signals to Search Engines" (<http://www.acronym.com/new-signals-to-search-engines.html>). He summarizes the state of web search as follows:

We're essentially trying to force elephants into browsers that do not want them. The browser that Sir Tim Berners Lee invented, along with HTML and the HTTP protocol, was intended to render text and graphics on a page delivered to your computer via a dial-up modem, not to watch movies like we do today. Search engine crawlers were developed to capture text from HTML pages and analyze links between pages, but with so much information outside the crawl, is it the right method for an always-on, ever-demanding audience of self producers?

Universal Search was a step that acknowledged part of this problem by making all types of data available through web search. But many of these data types do not provide the traditional text-based signals that search engines rely on. Here is more from Mike Grehan's paper:

Signals from end users who previously couldn't vote for content via links from web pages are now able to vote for content with their clicks, bookmarks, tags and ratings. These are very strong signals to search engines, and best of all, they do not rely on the elitism of one web site owner linking to another or the often mediocre crawl of a dumb bot. You can expect that these types of signals will become a significant factor in the future.

It provides the search engines with information on data types they cannot currently process (such as images and video), and it provides them with another tool to fight spam. This type of data already affects the rankings of videos on sites such as YouTube, which in January 2009 had become the Web's second largest search engine (according to comScore, Hitwise, and Nielsen Online). YouTube's ascent in search volume is particularly interesting because it is not a general web search engine, but one that focuses on a specific vertical, that of videos. This speaks to demand shifts taking place among the consumers of search results.

At the end of the day, the best results are likely to be provided by the best sites (there are exceptions; for example, for some search queries the best results may be "instant answers"). The technology the engines have now rewards a very select subset of web properties that have success with two ranking signals: good keyword targeting and good (or lots of moderate to crappy) links.

More data collection means more opportunity to win, even if your site doesn't conform flawlessly to these signals, and a better chance that if these are the only indicators you're winning, you could be in big trouble. Keywords and links will likely remain the primary ranking factors until 2012 or later, but the evolution of search engines in the direction of using the wisdom of the crowds is steadily gaining momentum and strength.

Following these advances, what will be next? Other areas that people are working on are artificial intelligence (AI) and linguistic user interfaces (LUIs). LUIs are voice-driven interfaces, whose arrival would completely transform the human-computer interface and how people search, work, and play. It will be far easier to talk to your computer and tell it what to do than trying to type those instructions in on a keyboard. According to an Acceleration Watch article at <http://www.accelerationwatch.com/lui.html>, you can plan to see these trends emerging between 2012 and 2019. Here are some excerpts from that article:

Clearly the keyboard is a primitive, first-generation interface to our personal computational machines. It gives us information, but not symbiosis. We humans do not twiddle our fingers at each other when we exchange information. We primarily talk, and use a rich repertoire of emotional and body language in simultaneous, noninterfering channels.

And:

In other words, talking is the highest, most natural, and most inclusive form of human communication, and soon our computers everywhere will allow us to interface with them in this new computational domain.

When these types of technologies will arrive is not something anyone can predict with certainty. Recent history has been littered with new technological advances that were supposedly on the verge of happening, but took much, much longer than predicted.

Google's Dominance

Thousands of posts, news articles, and analysis pieces have covered the central topic of battling Google's dominance in web search, but few have discussed the most telling example of the search giant's dominance. Many believe that the key to Google's success, and more importantly, a key component in its corporate culture, is its willingness and desire to get search users going to the destination site as quickly as possible.

Some also believe that Google's biggest barrier to entry in the search engine market is its advertising platform, which is the world's largest. By expanding its search, it is able to create a more enticing advertising platform through AdWords, AdSense, and its embeddable Google Search box.

However, it goes a bit deeper than that. In late 2008, tests were performed in which users were asked which search engine's results they preferred for a wide variety of queries—long tail searches, top-of-mind searches, topics about which their emotions ranged from great passion to total agnosticism. They were shown two sets of search results and were asked which they prefer (see [Figure 13-1](#)).

Lots of tests such as this have been run with all sorts of differentiations. In some, the brands are removed so that users see only the links. Testers do this to get an idea of whether they can win from a pure "quality" standpoint. In others, the brands remain to get an unvarnished and more "real-world" view. And in one particular experiment—performed many times by many different organizations—the results are swapped across the brands to test whether brand loyalty and brand preference are stronger than qualitative analysis in consumers.

It is this last test that has the most potentially intriguing results. Because in virtually every instance where qualitative differences weren't glaringly obvious, Google was picked as the best "search engine" without regard for the results themselves (see [Figure 13-2](#)).

Fundamentally, testers find (again and again) that the brand preference for Google outweighs the logical consideration of the search results quality.

Search engines that plan to take market share from Google are going to have to think differently. If Microsoft or Yahoo! or a start-up search engine wants to take market share, it's going to have to think less like a technology company trying to build a better mousetrap and

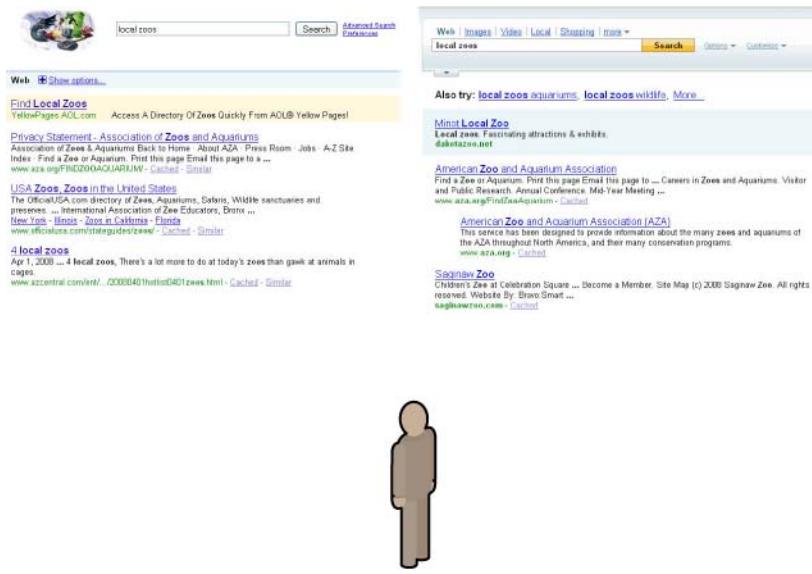


FIGURE 13-1. Comparing Google and Yahoo! results

more like a brand trying to win mind share from a beloved competitor. How did Pepsi take share away from Coke? Or Toyota from Ford? That is beyond the scope of this book, but it is a process that can take more than a great idea or great technology. It requires a massive psychological shift in the way people around the world perceive the Google brand against its competition.

Also consider the official Google mission statement: "Google's mission is to organize the world's information and make it universally accessible and useful." It is already moving beyond that mission. For example, Google and NASA are working on new networking protocols that can work with the long latency times and low bandwidth in space.

Google is also pursuing alternative energy initiatives (<http://googleblog.blogspot.com/2008/10/clean-energy-2030.html>), which clearly goes beyond its mission statement. In addition, Google has ventures in office productivity software with Google Docs (<http://docs.google.com>). These two initiatives have little to do with SEO, but they do speak to how Google is trying to expand its reach.

Another potential future involves Google becoming a more general-purpose pattern-matching and searching engine. The concept of performing pattern matching on text (e.g., the current Google on the current Web) is only the first stage of an evolving process. Imagine the impact if Google turns its attention to the human genome and creates a pattern-matching engine that revolutionizes the way in which new medicines are developed.



FIGURE 13-2. Results indicating that users may have a strong emotional preference for Google

More Searchable Content and Content Types

The emphasis throughout this book has been on providing the crawlers with textual content semantically marked up using HTML. However, the less accessible document types—such as multimedia, content behind forms, and scanned historical documents—are being integrated into the search engine results pages (SERPs) more and more, as search algorithms evolve in the ways that the data is collected, parsed, and interpreted. Greater demand, availability, and usage also fuel the trend.

Engines Will Make Crawling Improvements

The search engines are breaking down some of the traditional limitations on crawling. Content types that search engines could not previously crawl or interpret are being addressed. For example, in mid-2008 reports began to surface that Google was finding links within JavaScript (<http://www.seomoz.org/ugc/new-reality-google-follows-links-in-javascript-4930>). Certainly, there is the possibility that the search engines could begin to execute JavaScript to find the content which may be embedded within it.

In June 2008, Google announced that it was crawling and indexing Flash content (<http://googlewebmastercentral.blogspot.com/2008/06/improved-flash-indexing.html>). In particular, this announcement indicated that Google was finding text and links within the content. However, there were still major limitations in Google's ability to deal with Flash-based content. For

example, it applied only to Flash implementations that do not rely on external JavaScript calls, which is something that many Flash-based systems use.

Perhaps the bigger problem is the fact that Flash is not inherently textual. It is essentially like any other video where there is little incentive within the medium to use lots of text, and that limits what the search engine can interpret. So, although this is a step forward, the real returns for people who want to build all-Flash sites will probably need to wait until social signals become a stronger factor in search rankings.

Another major historical limitation of search engines is dealing with forms. The classic example is a search query box on a publisher's website. There is little point in the search engine punching in random search queries to see what results the search engines return. However, there are other cases in which a much simpler form is in use, such as a form that a user may fill out to get access to a downloadable article.

Search engines could potentially try to fill out such forms, perhaps according to a protocol where the rules are predefined to gain access to such content in a form where they can index it and include it in their search results. A lot of valuable content is currently isolated behind such simple forms, and defining such a protocol is certainly within the realm of possibility (though it is no easy task, to be sure). Google has stated that it has this capability, but will use it only on very important but inaccessible sites (<http://googlewebmastercentral.blogspot.com/2008/04/crawling-through-html-forms.html>).

This is but one specific example, and there may be other scenarios where the search engines might perform form submissions and gain access to currently inaccessible content.

Engines Are Getting New Content Sources

As we noted earlier, Google's stated mission is "to organize the world's information and make it universally accessible and useful." This is a powerful statement, particularly in light of the fact that so much information has not yet made its way online.

As part of its efforts to move more data to the Web, in 2004 Google launched an initiative to scan in books so that they could be incorporated into a Book Search (<http://books.google.com/>) search engine. This became the subject of a lawsuit by authors and libraries, but evidently a settlement was reached in late 2008 (<http://books.google.com/googlebooks/agreement/>). The agreement is still subject to full ratification by the parties, but that is expected to be resolved before the end of 2009. In addition to books, other historical documents are worth scanning. Google is not the only organization pursuing such missions (e.g., see <http://www.recaptcha.net>).

Similarly, content owners retain lots of other proprietary information that is not generally available to the public. Some of this information is locked up behind logins for subscription-based content. To provide such content owners an incentive to make that content searchable, Google came up with its First Click Free concept (discussed in [Chapter 6](#)), which is a program to allows Google to crawl subscription-based content.

However, a lot of other content out there is not on the Web at all, and this is information that the search engines want to index. To access it, they can approach the content owners and work on proprietary content deals, and this is also an activity that the search engines all pursue.

Multimedia Is Becoming Indexable

Content in images, audio, and video is currently not indexable by the search engines, but all the major engines are working on solutions to this problem. In the case of images, optical character recognition (OCR) technology has been around for decades. The main challenge in applying it in the area of search has been that it is a relatively compute-intensive process. As computing technology continues to get cheaper and cheaper, this becomes a less difficult problem.

In the meantime, creative solutions are being found. Google is already getting users to annotate images under the guise of a game, with Google Image Labeler (<http://images.google.com/imagelabeler/>). In this game, users agree to record labels for what is in an image. Participants work in pairs, and every time they get matching labels they score points, with more points being awarded for more detailed labels.

Or consider <http://recaptcha.net>. This site is helping to complete the digitization of books from the Internet Archive and old editions of the *New York Times*. These have been partially digitized using scanning and OCR software. OCR is not a perfect technology and there are many cases where the software cannot determine a word with 100% confidence. However, [Recaptcha.net](http://recaptcha.net) is assisting by using humans to figure out what these words are and feeding them back into the database of digitized documents.

First, Recaptcha.net takes the unresolved words and puts them into a database. These words are then fed to blogs that use the site's CAPTCHA solution for security purposes. These are the boxes you see on blogs and account sign-up screens where you need to enter the characters you see, such as the one shown in [Figure 13-3](#).



FIGURE 13-3. Recaptcha.net CAPTCHA screen

In this example, the user is expected to type in *morning*. However, in this case, Recaptcha.net is using the human input in these CAPTCHA screens to help it figure out what the word was in the book that was not resolved using OCR. It makes use of this CAPTCHA information to improve the quality of its digitized book.

Similarly, speech-to-text solutions can be applied to audio and video files to extract more data from them. This is also a relatively compute-intensive technology, so it has not yet been applied in search. But it is a solvable problem as well, and we should see search engines using it within the next decade.

The business problem the search engines face is that the demand for information and content in these challenging-to-index formats is increasing exponentially. Search results that do not include this type of data, and accurately so, will begin to be deemed irrelevant or wrong.

The emergence of YouTube in late 2008 as the #2 search engine (ahead of Yahoo! and Microsoft) is a powerful warning signal. Users want this alternative type of content, and they want a lot of it. User demand for alternative forms of content will ultimately rule the day, and they will get what they want. For this reason, the work on improved techniques for indexing such alternative content types is an urgent priority for the search engines.

Interactive content is also growing on the Web, with technologies such as Flash and AJAX leading the way. In spite of the indexing challenges these technologies bring to search engines, the use of these technologies is continuing because of the experience they offer for users who have broadband connectivity. The search engines are hard at work on solutions to better understand the content wrapped up in these technologies as well.

Over time, our view of what is “interactive” will change drastically. Two- or three-dimensional first-person shooter games and movies will continue to morph and become increasingly interactive. Further in the future, these may become full immersion experiences, similar to the Holodeck on “Star Trek.” You can also expect to see interactive movies where the audience influences the plot with both virtual and human actors performing live. These types of advances are not the immediate concern of today’s SEO practitioner, but staying in tune with where things are headed over time can provide a valuable perspective.

Search Becoming More Personalized and User-Influenced

Personalization efforts have been underway at the search engines for some time. As we discussed in [Chapter 2](#), the most basic form of personalization is to perform a reverse IP lookup to determine where the searcher is located, and tweak the results based on the searcher’s location. However, the search engines continue to explore additional ways to expand on this simple concept to deliver better results for each user. It is not yet clear whether personalization has given the engines that have invested in it heavily (namely Google) better results overall or greater user satisfaction, but their continued use of the technology suggests that, at the least, their internal user satisfaction tests have been positive.

Determining User Intent

The success of Internet search has always relied (and will continue to rely) on search engines’ abilities to identify searcher intent. Microsoft has branded [Bing.com](#), its latest search project,

not as a search engine but as a “decision” engine. It chose this label because of what it found in its research and analysis of search sessions. The slide shown in [Figure 13-4](#) was presented by Satya Nadella at the Microsoft Search Summit 2009 in June 2009.

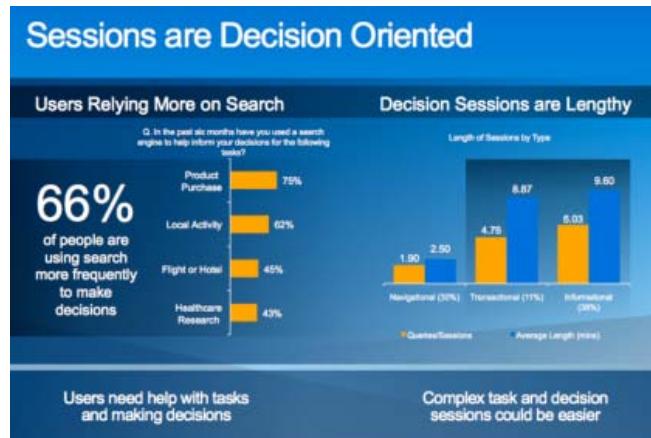


FIGURE 13-4. Microsoft analysis of search sessions

The conclusion was that about two-thirds of searchers frequently use search to make decisions. Microsoft also saw that making these decisions was proving to be hard based on the average length of a search session. What makes this complex is that there are so many different modes that a searcher may be in. Are searchers looking to buy, to research, or just to be entertained? Each of these modes may dictate very different results for the same search.

Google personalization and Universal Search are trying to tap into that intent as well, based on previous search history as well as by serving up a mix of content types, including maps, blog posts, videos, and traditional textual results. Danny Sullivan, editor-in-chief of *Search Engine Land*, added to the discussion on the importance of relevancy in how the information is presented, such as providing maps for appropriate location searches or the ability to list blog results based on recency as well as relevancy. It is not just about presenting the results, but about presenting them in the format that matches the searcher’s intent.

It could be as easy as letting the user reveal her intent. The now-defunct Yahoo! Labs project Yahoo! Mindset simply had a searcher-operated slider bar with “research” on one end and “buy” on the other. Sliding it reshuffled the results in real time via AJAX.

User Interactions

One area that will see great exploration will be in how users interact with search engines. As RSS adoption continues to grow and the sheer amount of information in its many formats expands, users will continue to look to search engines to be not just a search destination, but

also a source of information aggregation; the search engine as portal, pulling and updating news and other content based on the user's preferences.

Marissa Mayer, Google's VP of Search User Experience and Interface Design, made a particularly interesting comment that furthers the sense that search engines will continue their evolution beyond search:

I think that people will be annotating search results pages and web pages a lot. They're going to be rating them, they're going to be reviewing them. They're going to be marking them up...

The separate mention of "web pages" may be another reason why the release of Google Chrome (<http://www.google.com/chrome>) was so important. Tapping into the web browser might lead to that ability to annotate and rate those pages and further help Google identify what content interests the user.

Although Chris Sherman, executive editor of *Search Engine Land*, feels that advancement within search personalization is still fairly limited, he offered up an interesting interactive approach that the search engines might pursue, as a way to allow users to interact with search engines and help bring about better results:

Until search engines can find a way let us search by example - submitting a page of content and analyzing the full text of that page and then tying that in conjunction with our past behavior...

New Search Patterns

This is all part of downgrading those things that have been synonymous with SEO, such as rankings, keywords, and optimization, to a much greater focus on the users, tying into their intent and interests at the time of search. Personalization will make site stickiness ever more important. Securing a position in users' history, becoming an authoritative go-to source for information, will be more critical than ever. Winning in the SERPs will require much more than just position.

Over time, smart marketers will recognize that the attention of a potential customer is a scarce and limited quantity. As the quantity of information available to us grows, the amount of time we have available for each piece of information declines, creating an attention deficit. How people search, and how advertisers interact with them, may change dramatically as a result.

In July 2008, *The Atlantic* cover article was titled "Is Google Making Us Stupid?" (<http://www.theatlantic.com/doc/200807/google>). The thrust of this article was that Google was so powerful in its capabilities that humans need to do less (and less!). This is a trend that will continue. After all, who needs memory when you have your "lifestream" captured 24/7 with instant retrieval via something akin to "Google Desktop Search"? When you have instant perfect recall of all of human history?

These types of changes, if and when they occur, could transform what we today call "SEO" into something else, where the SEO of tomorrow remains responsible for helping publishers

gain access to potential customers through a vast array of new mechanisms that currently do not exist.

User-Driven Search Results

In November 2008 Google announced SearchWiki (<http://googleblog.blogspot.com/2008/11/searchwiki-make-search-your-own.html>). This was a move by Google to allow individual end users to tweak their search results. You can see the controls on the search results screen shown in Figure 13-5.

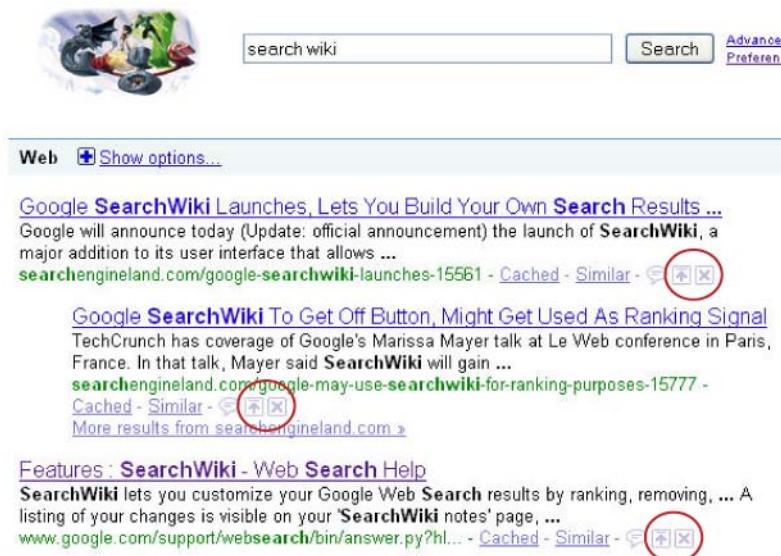


FIGURE 13-5. Google's SearchWiki

Users must be logged in to use SearchWiki, but if they are, they can move results up, move them back down again, or erase them from the results altogether. SearchWiki will remember their stated preferences in later sessions. There are many other features as well, such as the ability for users to comment on specific search results.

Although Google stated initially that it was not using SearchWiki to drive rankings, it has since clarified that it plans to use that data in the future (<http://www.stonetemple.com/articles/interview-cedric-dupont.shtml>). SearchWiki will potentially provide Google with a wealth of data directly from end users on what they like and do not like.

Google extended this basic concept in January 2009 with the notion of Preferred Sites. The Google Operating System blog broke the news (<http://googlesystem.blogspot.com/2009/01/google-preferred-sites.html>), and users who were included in the test program by Google are able to

specify sites that they consider authoritative. Those sites are then given greater weight in the results for those users.

This, of course, could be yet another powerful ranking signal for the search engines. The initial SearchWiki announcement was about how users could tailor the results for specific queries. Voting a site up on one query would not affect its results on other search queries.

Now, with Preferred Sites, the user can make a broader sweeping decision that she wants a particular site to be emphasized whenever she enters a search query for which the selected site has a relevant result. A paid subscriber to Encyclopedia Britannica (<http://www.britannica.com>) could prefer that site over Wikipedia and, using Preferred Sites, can ensure that Wikipedia doesn't appear at the top of search results, while Britannica does.

Growing Reliance on the Cloud

Cloud computing is transforming how the Internet-connected population use computers. Oracle founder Larry Ellison's vision of thin-client computing may yet come to pass, but in the form of a pervasive Google OS "operating system" and its associated, extensive suite of applications. Widespread adoption by users of cloud-based (rather than desktop) software and seemingly limitless data storage, all supplied for free by Google, will usher in a new era of personalized advertising within these apps.

Google is actively advancing the mass migration of desktop computing to the cloud, with initiatives such as Google Docs & Spreadsheets (<http://docs.google.com>), Gmail (<http://mail.google.com>), Google Calendar (<http://calendar.google.com>), Google Reader (<http://google.com/reader>), and the much anticipated Google Wave and Google Web Drive/Gdrive—all with Google Chrome as the wrapper. These types of services lead to users entrusting their valuable data to the Google cloud. This brings them many benefits (but also concerns around privacy, security, uptime, and data integrity).

For example, most users don't do a good job of backing up their data, making them susceptible to data loss from hard drive crashes and virus infections. In addition, they can more easily access that information from multiple computers (e.g., their work and home computers).

Google benefits by having a repository of user data available for analysis—which is very helpful in Google's quest to deliver ever more relevant ads and search results. It also provides multiple additional platforms within which to serve advertising. Furthermore, regular users of a service such as Google Docs are more likely to be logged in a greater percentage of the time when they are on their computers, which is important because Google serves personalized results only when users are logged in.

The inevitable advance of cloud computing will offer more and more services with unrivaled convenience and cost benefits, compelling users to turn to the cloud for their data and their apps. Data portability from and to these services will become a critical issue. [OpenID](#) and [Open](#)

Social are two of the early data portability solutions leading the charge, which will also help shepherd in this era of cloud computing.

Increasing Importance of Local, Mobile, and Voice Recognition Search

New forms of vertical search are becoming increasingly important. Areas that the search engines have already embraced include local search and mobile search. In addition, voice-recognition-based search is an area in which all the major engines are actively investing.

Local Search

Local search was an active business before the advent of the Internet, when Yellow Pages as supplied by your phone company were the tool of choice. As the World Wide Web first became prominent in our lives, the first way that this unfolded on the Web was in the form of Yellow Pages migrating online. These types of sites have themselves evolved to leverage more of the unique nature of the Web. Some of the major players are [MapQuest](#), [CitySearch](#), [Local.com](#), and [Superpages](#). You can see the relative market share of these four sites in [Figure 13-6](#), which depicts a chart that [Compete.com](#) created in March 2009.

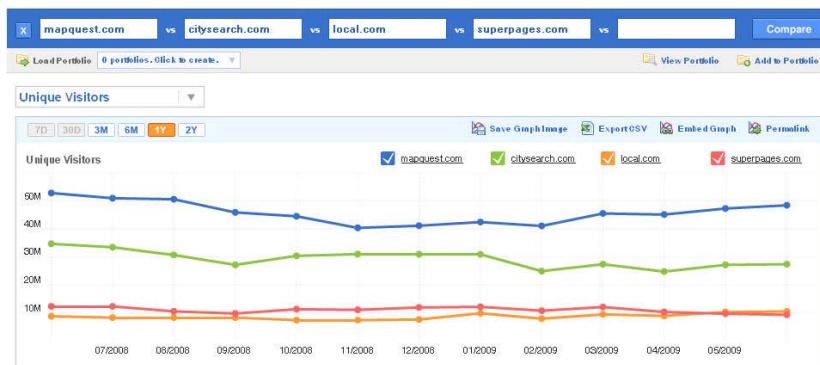


FIGURE 13-6. Major local search websites

As the major search engines evolved, they integrated sophisticated mapping technology into their systems, and then began to map in local business data as well. You can find these local search engines at <http://maps.google.com/>, <http://local.yahoo.com>, and <http://www.bing.com/maps>. These search engines map in some of their business data from the same types of data sources that drive the Yellow Pages websites, such as [Axiom](#), [InfoUSA](#), and [Localeze](#).

Still, the more traditional sites are doing reasonably well in terms of market share, as shown in the Complete.com chart in [Figure 13-7](#).

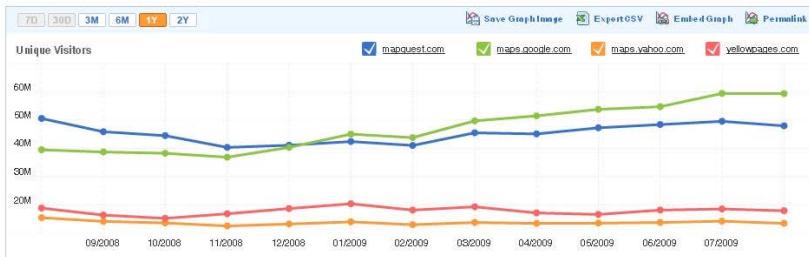


FIGURE 13-7. MapQuest still ahead of local search engines

Figure 13-7 shows that MapQuest and Yellowpages.com are still drawing significant traffic. Keeping in mind our earlier chart on the traffic at the major search portals, you can see that CitySearch, Superpages, and Local.com would have material traffic and placement on this chart as well. Note the steady increase in Google's market share, and the arrows that highlight that Google's growth comes at the expense of the local portals.

The next major advent in the realm of local search was initiated when Google announced Universal Search in May 2008. As a result of this announcement, you could get local search results without having to go to <http://local.google.com>, and these results began to appear prominently in the basic web search results. Figure 13-8 shows an example search for *boston rental cars*.

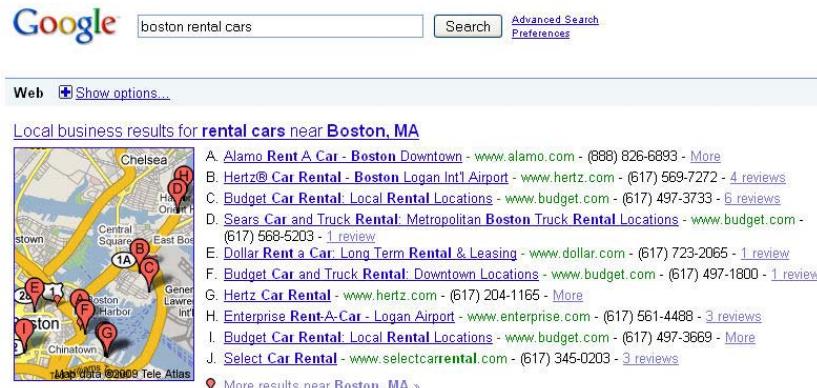


FIGURE 13-8. Search results for “boston rental cars”

As is common with these types of queries, the local results show up at the top, under the ads, and before the regular web search results. This is prime positioning, and this increased the intensity with which publishers began to focus on local SEO (a topic we covered at length in Chapter 8).

One reason for all these moves is that local search volume has skyrocketed. Just how much of search has local intent? Greg Sterling was able to get some research work performed for him by comScore that suggested that 35% to 40% of search queries had some sort of local search intent (<http://gesterling.wordpress.com/2007/03/01/local-numbers-setting-the-record-straighter/>).

One of the big challenges facing the local search engines is *map spam*, the proliferation of spam results in local search. Here are URLs to two articles from October 2008 that provide numerous examples:

- <http://searchengineland.com/google-local-business-hijacking-microsoft-acquires-yahoo-becomes-escort-service-15313>
- <http://searchengineland.com/problems-continue-with-google-local-business-listings-15087>

Resolving these issues is obviously critical for the search engines and is a major area of investment for them. We believe that local search is an area whose importance will continue to grow. With a very large number of searches having local intent, this is potentially a major frontier for developments in search.

Over time, you can expect that these problems will be worked out. Local search is already extremely powerful, but you can anticipate that the search engines will control the flow of customers in cyberspace and the real world, because customers will be finding merchants through search. Success in local search will be what makes or breaks many businesses.

Consumers are becoming increasingly reliant on local search and its auxiliary services—street maps, directions, satellite views, street views, 3D visualizations (Google Earth), enhanced listings, user reviews, and ratings. These tools guide consumers' movements and their decision-making processes; this means the search engines are in a powerful position to manage the flow of customers from cyberspace to brick-and-mortar businesses. So, in a way, the search engines are in the logistics business, building the connection between online and offline.

Mobile Search

Mobile search has been a hot topic in the industry for a few years now, with frequent predictions that “this year will be the year of mobile search”. Although this has not happened in quite the way people have expected, mobile Internet usage and mobile search have seen steady growth. The increased use of smartphones has proven to be a driving factor in this growth, with Asia and Europe leading the way.

U.S. market

The U.S. market is gradually catching up, though. In November 2008, Marketing Charts reported the following (<http://www.marketingvox.com/mobile-local-search-041827/>):

A rapid rise in the rate of US smartphone adoption is driving a corresponding increase in local mobile search activity, according to the latest wave of the “Mobile Market View” (<http://www.kelseygroup.com/research/mobile-market-view.asp>), a consumer study of US mobile phone users from The Kelsey Group (<http://www.kelseygroup.com/>) and ConStat (<http://www.constat.com>).

Marketing Charts also noted that 19% of U.S. mobile users were using smartphones, and 49% planned to buy them within the next two years. The 49% figure comes from a survey of consumers by The Kelsey Group.

Similarly, The Kelsey Group found that mobile usage of the Internet is growing at a significant clip in the United States, where between September 2007 and October 2008 mobile Internet usage increased 20%. How much of this is search? Additional data from The Kelsey Group reveals that users performing Internet searches grew from 9.8% to 15.6% of mobile users. That is nearly 60% growth in 13 months. This data shows us that rapid growth is in progress right now, and it is reasonable to conjecture that the growth will continue.

Over time, you can expect the world to become untethered. The Internet of tomorrow is a mobile Internet. This trend is even more evident in developing nations. According to mobiForge (<http://mobiforge.com/analysts/blog/mobile-internet-users-india-outnumber-pc-internet-users-3-1>) in India, mobile Internet users outnumber PC Internet users three to one. In one year, the number of mobile Internet users in India doubled to account for 9% of mobile Internet users worldwide. And according to The Hindu (<http://www.hindu.com/thehindu/holnus/002200706070350.htm>), one in 11 of mobile Internet users worldwide are in India.

Even in the United States, the convenience of being able to get the information you need while on the go is just way too compelling. Why be tied to a desk if you do not have to be? Further, in many countries, freedom from a desktop is a requirement because the space that people have for their personal lives is quite different from what people experience in the United States.

The ever-expanding versatility and power of mobile devices—from indispensable utility apps, to immersive multimedia players, to massively multiplayer online games, to paradigm-shifting hardware advances such as the iPhone’s multitouch display, proximity sensor, GPS, and gyroscopes—will fuel this growth. The *network effect*, whereby the value of the network grows by the square of the size of the network (Metcalf’s Law), gives further incentive for users to migrate to their mobile devices as more and more apps allow them to interact with their peers in increasingly more interesting ways.

The small keyboard/typing surface is currently a severe limitation, but the search engines are working on voice-based solutions to circumvent this. The advent of the LUI will certainly revolutionize mobile search, and when this happens a whole new set of skills will be required of the SEO practitioner.

Worldwide mobile Internet growth

As we noted in “Country-Specific Search Engines” on page 74 in Chapter 2, China has more than 600 million mobile phone users, which is approaching double the entire U.S. population.

Currently, only about 12% of those users (72 million) are accessing the Internet with their phones, largely because the telephone carriers there do not have 3G (high-speed Internet for use over the phone system) licenses at the moment.

In March 2009, Interfax released a report (<http://www.interfax.com/4/479860/news.aspx>) that indicated that China's mobile Internet industry generated more than \$650 million in revenue in 2008, up 73.2% from 2007, according to a research report released in March 2009. Mobile search is still a small component of the revenue, though, with the report indicating that only 3.8% of the revenue comes from it.

CCID analyst Wei Guichuan indicated that "the low speed and high cost of mobile Internet services in China has held back the sector's growth," but the report also said that there will be 240 million mobile Internet service users in China in 2009.

Looking to Japan, a September 2007 article in *The Guardian* (<http://www.guardian.co.uk/technology/2007/sep/27/guardianweeklytechnologysection.mobilephones>) reported that in a nation of 127 million, there were more than 100 million mobile Internet subscribers. When you consider the lack of utility of mobile phones to infants and small children, this is pretty close to 100% saturation. The broad acceptance of mobile devices in Asia is not limited to China and Japan, either. Many other Asian countries have significant numbers of mobile device users, such as Korea, Taiwan, India, and more.

Western Europe is also very advanced in its acceptance of mobile devices and the mobile Internet. According to a report by Forrester Research (<http://www.forrester.com/Research/Document/Excerpt/0,7211,42199,00.html>), "by the end of 2013, one-quarter of mobile phone users will have a 3.5G-capable device. By that time, we expect 38% of mobile subscribers to use mobile Internet services at least once per month."

According to the African Mobile Phone Fact Book (http://www.web4dev.org/images/8/8d/Africa_Mobile_Fact_Book_2008.pdf), growth of mobile device usage in Africa is also projected to be rapid. The year 2008 ended with about 370 million mobile subscribers and the report projects that there will be more than 560 million mobile subscribers by the end of 2012. The use of 3G devices is still quite low, but almost 19% of handsets are expected to be 3G-capable by the end of 2011.

By all counts, the growth is rapid and is a natural consequence of the attractiveness of the mobile device. Being chained to your desk to access the Internet is not a problem when you are working on major tasks. But if you are out and about and you need some quick information, it is a huge value to be able to pull out your mobile device and get it.

What remains unresolved is the revenue model that will drive mobile search services. Presumably, mobile search revenue will be advertising-based, but it is not clear in the mobile environment that people will be that accepting of ads. After all, they may not be in the mindset to buy something when they are mobile. When the LUI becomes the primary interface, desktop computing will be pervasive. "Calling" your computer for a project discussion will likely be as common as calling the office.

One thing that is clear, though, is that the ability to conduct searches when you need it on a mobile device will become a requirement, so that means some mechanism must come into being to help pay for it.

U.S. mobile search market share

Google is the dominant player in mobile search in the United States (per Nielsen Mobile, <http://www.nielsenmobile.com/html/press%20releases/GoogleandYahooSearchData.html>). Figure 13-9 shows a summary of the company's data for Q1 2008.

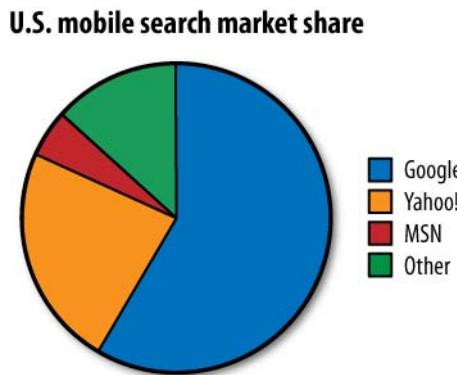


FIGURE 13-9. Mobile search market share most popular mobile features

So, there is nothing new here in terms of Google's dominance. Nonetheless, mobile search is a new environment, and new environments always create the potential for change.

Voice Recognition Search

When users are mobile they must deal with the limitations of their mobile device, specifically the small screen and small keyboard. These make web surfing and mobile searching more challenging than they are in the PC environment.

Voice search could be a great way to improve the mobile search experience. It eliminates the need for the keyboard, and provides users with a simple and elegant interface. Speech recognition technology has been around for a long time, and the main challenge has always been that it requires a lot of computing power.

Two examples already exist:

GOOG-411

This is a free service from Google that you call at (800) GOOG-411. The system is voice-activated, and you can use it to find businesses you are looking for and get maps sent to your phone.

Google Mobile App for iPhone

Included in this application is voice searching capability. You can speak your query into the application and the results are displayed on-screen.

Processing power continues to increase, even on mobile devices, and the feasibility of this type of technology is growing. This should be another major area of change in the mobile search landscape.

In addition, it is reasonable to expect that voice recognition technology will be applied to the actual recorded content of audio and video files, to determine content and aid in ranking these media types within search results, something that will likely lead to video “script” optimization as an added component of video SEO.

Increased Market Saturation and Competition

One thing you can count on with the Web is continued growth. In a July 2008 post, Google indicated that it was aware of 1 trillion web pages (<http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>). That's a lot of pages! Considering that the known count of pages in 2000 was about 1 billion, this represents a growth rate of 1,000 times in eight years.

We should note that a lot of these pages may be low-quality or duplicate-content-type pages that will never see the light of day. For example, at the SES New York conference in 2009, Sharad Verma of Yahoo! estimated that about one-third of the pages on the Web were actually duplicates of other pages. In addition, the same Google post noted that Google has “only” 40 billion pages indexed. No matter how you count it, the Web is a big place, but one where the signal-to-noise ratio is very low.

One major trend emerges from an analysis of Internet usage statistics. According to Miniwatts Marketing Group, 74.4% of the North American population uses the Internet (<http://www.internetworldstats.com/stats.htm>), so there is not much room for growth there. In contrast, Asia, which already has the most Internet users with 657 million, has a penetration rate of only 17.4%. Other regions with a great deal of opportunity to grow are Africa, the Middle East, and Latin America.

This data is telling us that in terms of the number of users, North America is already approaching saturation. Europe has some room to grow, but not that much. However, in Asia, you could see two times to four times that growth, or 1.5 billion to 2 billion users! The bottom line is that a lot of Internet growth in the coming decade will be outside North America and that will provide unique new business opportunities for those who are ready to capitalize on that growth.

With this growth has come an increasing awareness of what needs to be done to obtain traffic. The search engines are the dominant suppliers of traffic for many publishers, and will continue to be for some time to come. For that reason, awareness of SEO will continue to increase over time. Here are some reasons why this growth has continued:

The Web outperforms other sales channels

When organizations look at the paths leading to sales and income (a critical analysis whenever budgets are under scrutiny), the Web almost always comes out with one of two assessments. Either it is a leading sales channel (especially from an ROI perspective) or it is deemed to be an area with the greatest opportunity for growth. In both scenarios, web marketing and, in correlation, SEO, takes center stage.

It is the right time to retool

Established companies frequently use down cycles as a chance to focus attention inward and analyze themselves. Consequently, there's a spike in website redesigns and SEO along with it.

Paid search drives interest in SEO

Paid search spending is still reaching all-time highs, and when companies evaluate the cost and value, there's a nagging little voice saying "70%+ of the clicks do not even happen in the ads; use SEO." Data from Enquisite suggests that the cost of the SEO effort may be as little as one-seventh the cost of PPC as well (<http://www.enquisite.com>).

SEO is losing its stigma

Google is releasing SEO guides, Microsoft and Yahoo! have in-house SEO departments, and the "SEO is BS" crowd have lost a little of their swagger and a lot of their arguments. No surprise—solid evidence trumps wishful thinking, especially when times are tough.

Marketing departments are in a brainstorming cycle

A high percentage of companies are asking the big questions: "how do we get new customers?" and "what avenues still offer opportunity?" Whenever that happens, SEO is bound to show up near the top of the "to be investigated" pile.

Search traffic will be relatively unscathed by the market

Sales might drop, conversion rates might falter a bit, but raw search traffic isn't going anywhere. A recession doesn't mean people stop searching the Web, and with broadband adoption rates, Internet penetration, and searches per user consistently rising, search is no fad. It is here for the long haul.

Web budgets are being reassessed

We've all seen the news about display advertising falling considerably; that can happen only when managers meet to discuss how to address budget concerns. Get 10 Internet marketing managers into rooms with their teams and at least four or five are bound to discuss SEO and how they can grab that "free" traffic.

Someone finally looked at the web analytics

It is sad, but true. When a downturn arrives or panic sets in, someone, maybe the first someone in a long time, checks the web analytics to see where revenue is still coming in. Not surprisingly, search engine referrals with their exceptional targeting and intent matching are ranking high on the list.

Although more and more people are becoming aware of these advantages of SEO, there still remains an imbalance between paid search and SEO. The following data is from SEMPO on the allocation of search marketing budgets:

- In 2004 it was 85% PPC versus 12% organic.
- In 2005 it was 87% PPC versus 11% organic.
- In 2006 it was 87% PPC versus 12% organic.
- In 2007 it was 88% PPC versus 10% organic.
- In 2008 it was 88% PPC versus 10% organic.

This is in spite of the fact that for every click on a paid search result roughly seven to eight clicks take place on organic search results (according to [Enquisite](#)). Additional data from Enquisite indicates that paid search clicks do convert better than organic clicks, but only about 1.5 times better.

This suggests that the opportunity related to organic search is 5.6 times greater than the opportunity for paid search. This is an impressive opportunity gap, yet the SEMPO data presented earlier shows that the PPC share of search marketing spend grew from 2004 to 2008 (albeit only slightly).

This suggests that either SEO could see some growth to align budgets with potential opportunity, or firms that focus solely on SEO services had better diversify. We believe it likely that SEO budgets will begin to expand as more and more businesses better understand the mechanics of the Web. In the short term, PPC is easier for many businesses to understand, because it has more in common with traditional forms of marketing. Ultimately, though, SEO is where the most money can be found, and the dollars will follow once people understand that.

SEO As an Enduring Art Form

Today, SEO can be fairly easily categorized as having four major objectives:

- Make pages accessible to search engine crawlers.
- Find the keywords that searchers employ (understand your target audience) and make your site speak their language.
- Build content that users will find useful, valuable, and worthy of sharing. Ensure that they'll have a good experience on your site to improve the likelihood that you'll earn links and references.
- Earn editorial links from good sources by building inviting, link-worthy content and applying classic marketing techniques to the online world.

Note, though, that the tactics an SEO practitioner might use to get links from editorial sources have been subject to rapid evolution, and will continue to be. In addition, mastery of social

media environments is also likely to be an expertise that most SEO professionals will be required to have.

One thing that you can be sure about in the world of search is change, as forces from all over the Web are impacting search in a dramatic way.

To be an artist, the SEO practitioner needs to see the landscape of possibilities for her website, and pick the best possible path to success. This currently includes social media optimization expertise, local search expertise, video optimization expertise, an understanding of what is coming in mobile, and more. That's a far cry from the backroom geek of the late 1990s.

No one can predict what the future will bring and what will be needed to successfully market businesses and other organizations on the Web in two years, let alone 5 or 10. However, you can be certain that websites are here to stay for a long time, and that websites are never finished and need continuous optimization just like any other direct marketing channel. SEO expertise will be needed for a long time—and no existing person is better suited to map the changing environment and lead companies to success in this new, ever-evolving landscape than today's SEO practitioner.

Conclusion

SEO is both art and science. The artistic aspect of SEO requires dynamic creativity and intuition. The search engine algorithms are too complex to reverse-engineer every aspect of them. The scientific aspect involves challenging assumptions, testing hypotheses, making observations, drawing conclusions, and achieving reproducible results. These two ways of thinking will remain a requirement as SEO evolves into the future.

In this chapter, we conveyed some sense of what is coming in the world of technology, and in particular, search. Although the past decade has seen an enormous amount of change, the reality is that it has simply been the tip of the iceberg. There's a lot more change to come, and at an ever-increasing (exponential) rate. If the Law of Accelerating Returns holds, we're in for a wild ride.

In this fast-moving industry, the successful SEO professional has to play the role of early adopter. The early adopter is always trying new things—tools, tactics, approaches, processes, technologies—to keep pace with the ever-evolving engines. At the other end of the adoption curve are the laggards. Every industry has its laggards. In SEO, the laggards are still touting the benefits of meta keywords to their unsuspecting victims (er, customers).

It is not enough to adapt to change. You will need to embrace it, to evangelize it. Many in your (or your client's) organization may fear change, and steering them through these turbulent waters will require strong leadership. Thus, the successful SEO professional also has to play the role of change agent.

The need for organizations to find new customers will not diminish anytime soon, and neither will the need for searchable, web-based, and instantaneous access to information, products, and services. This ability—to connect the seeker with the provider—is perhaps the most valuable skill set on the Web today.

And although there is the possibility that the search engines could eventually be called “decision,” “dilemma,” or even “desire” engines, the absolute need for the understanding of and interactions between both the psychological and the technological natures of search will ensure that SEO as a discipline, and SEO professionals, are here to stay.

INDEX

Symbols

- " " (quotes, double), exact key phrase search, 54
- * (asterisk), wildcard search, 55
- + (plus sign) search operator, 54
- .. (dots), number range operator, 64
- 301 HTTP status code, 108, 255, 440
- 302 HTTP status code, 102, 255
- 404 errors reports, 416
- ~ (tilde) search operator, 44, 56
- (minus sign) search operator, 53

A

- A/B tests, 170
- absolute links, 233
- action tracking in analytics, 384–387
 - by referral source, 394
- Actionable Key Performance Indicators (KPIs), 375
 - web analytics data and, 387
- Axiom, 339, 546
- adCenter Keyword Generation Tool (Microsoft), 149
- Adobe Dreamweaver, 513
- Advanced Dork (Firefox plug-in), 59
- advertising purchases, direct link, 306
- AdWords accounts, 148
- AdWords Keyword Tool, 142
- affiliates
 - caution in republishing content, 225
 - thin affiliate sites, 462
 - thin affiliate sites filtered out by search engines, 232
- age differences in searching, 87
- age of sites, 50
 - advanced search technique for, 64
- aggregators, 339
- AJAX, 275
 - # delimiter, 199
- human-readable content invisible to search engines, 41
- problems for search engine crawlers, 198
- Alchemist Media, 196
 - anchor text example, 285
 - image tag example, 39
 - SEO blog, 501
- Alexa, 121
 - traffic comparison, 427, 428
- algorithmic ranking criteria, 32
- allinanchor: operator, 55
- allintitle: operator, 55
- allinurl: operator, 54
- Allow directive (robots.txt), 245
- alt attribute for images, 38, 67, 198
 - correct syntax for, 218
 - SEO audit of, 110
- Amacai, 339
- ambiguous queries, 46
- analytics (see web analytics)
- anchor text, 46, 105
 - customizable, on navigation links, 267
 - customizing on internal links for search optimization, 264
 - of inbound links, 50
 - influencing search engine rankings, 285
 - internal, keyword cannibalization and, 112
 - optimizing for blogs, 359
- Apache web servers
 - mod_rewrite module, 440
 - rewriting URLs using mod_rewrite, 257–261
 - URL redirecting, 256
- architecture, website (see information architecture)
- article directories, 314
- AskJeeves, 286
- attributions of sales, 389
- audience, understanding
 - importance of content, 86
 - segmenting site's audience, 87
 - thorough understanding of your products and services, 85

We'd like to hear your suggestions for improving our indexes. Send email to index@oreilly.com.

audio, difficulties for search engine crawlers, 197
 audits (site audit for SEO problems), 106–116
 elements of an audit, 106
 fixing internal linking problems, 113
 keyword cannibalization, 111
 keyword reviews, 110
 server and hosting issues, 115
 author profile pages, 360
 authority of links, 286
 source TLDs and, 293
 trust versus, 287
 automatically generated content, 123
 (see also CMSs)
 keyword targeting in, 219
 AWStats, 117, 377

B

B2B sites, action tracking on, 386
 backlinks
 checking with Bing Webmaster Tools, 495
 checking with Yahoo! advanced search operator, 59
 finding dubious links and link schemes, 118
 Firefox extensions for Yahoo! search operators, 61
 tools for checking, 109, 399
 Bing
 advanced search operators, 63
 DMCA notices to, 469
 Guidelines for Successful Indexing, 456
 indexing data from, 124
 Local Listing Center, 340
 reinclusion requests, 466
 search results layout, 24
 site indexing data, 398
 updating your Sitemap, 186
 vertical results in SERPs, 28
 vertical search, 66
 Celebrity xRank, 70
 image searches, 66
 WebLog, 474
 Webmaster Tools, 117, 493–497
 checking backlinks, 400
 Crawl Issues page, 416
 site indexing data, 398
 Yahoo! search to be powered by, 60
 Blended Search, 15, 71, 335
 Blendtec video campaign, 312, 326
 Blinkx, 369
 blog platforms, 357
 blog search, 69
 counts of, 138
 blogging for links, 298, 322
 Bloglines, 422
 blogosphere, tracking, 419

blogs
 action tracking on, 386
 duplicate content issues, 233
 leaving automatically generated comments on, 308
 linking to, 305
 local business, adding blog map or feedmap, 346
 NoFollow on links, 309
 optimizing for vertical search, 357–363
 anchor text, 359
 author profile pages, 360
 creating a great blog, 363
 sticky posts, 359
 structural optimizations, 358
 search engine blogs, 474
 SEO industry, 500–502
 tracking, 421–424
 links, 423
 subscribers, 421
 boldface text, 218
 Book Search (Google), 70
 Boolean searches, 43
 brand keywords versus non-brand keywords, 435
 branding
 considerations in planning stages of SEO, 100
 considerations when using SEO for, 92
 SEO for, 80
 using YouTube, 326
 browsers, 59
 (see also Firefox)
 AJAX # delimiter and, 199
 not supporting JavaScript, 39
 viewing source code, 35
 business assets
 combining with historical data for SWOT analysis, 132
 leveraging for SEO, 131
 business case for SEO, using analytics as, 376
 business factors affecting SEO plan, 84
 business listings in U.S., aggregators for, 339
 business planning, advanced methods, 88
 SMART objectives, 89
 SWOT analysis, 88
 business profiles, local, 342
 buying links, 302
 dangers of, 302
 detection by search engines, 320
 methods for, 306
 reporting of paid links to Google, 320
 strategies not considered as, 307
 buzz and public relations campaigns, tracking, 421

C

cache: operator, 56, 65

- CAN-SPAM Act, 304
- canonical content, 186, 187
- canonical redirects, 107, 260
- canonical URL tags, 233, 250
- CAPTCHAs, 254
- card sorting, 193
- category structuring of site subject matter, 192, 265
- CC (Creative Commons) licensed content, 65
- Celebrity xRank (Bing), 70
- chambers of commerce, 341
- charitable donations as method of buying links, 307
- charitable endeavors, link building through, 318
- China
 - Baidu search engine, 75
 - Internet usage, 76
 - mobile phone users, 550
- citation, 32
- citation analysis, 32
- CitySearch, 546
- classified sites, 341
- click tracking
 - for RSS feeds, 356
 - search results, natural versus paid, 16
 - distribution of search results and traffic, 17
 - intents and, 18
 - interaction, natural and paid search, 20
 - other factors, 20
- click-through rate (CTR), 150
- cloaking, 234
 - cookie and session-restricted pages and, 236
 - examples of websites performing, 239
 - by IP address range, 252
 - and segmenting content delivery, 238
 - by user agent, 252
- cloud computing, 545
- Clusty, related terms clusters, 139
- CMSs (content management systems), 263–268
 - duplicate content created by, 233
 - keyword targeting in, 219
 - selecting a CMS, 267
 - SEO issues, correcting, 263
 - third-party CMS add-ons, 268
 - unintentional creation of hidden text, 455
- code to text ratio, 223
- code, auditing for SEO problems, 110
- .com TLD, when to use another TLD, 206
- commentary from search engine employees, 474
- common search phrases, finding, 140
- community efforts assisting in online promotion, 346
- Compete, 121
 - traffic comparison, 428
 - traffic measurements, 427
- competition
 - competitors reporting you for spam, 461
 - determining top competitors, 118–122
 - analyzing their successful tactics, 120
 - finding the best, 119
 - spam examples, 118
 - finding out where competitors get links, 310
 - monitoring from SEO perspective, 397
 - understanding from SEO perspective, 100
 - use of DMCA to silence competitors, 469
- competitive analysis, 482–486
 - on a big budget, 486
 - blog tracking for, 421
 - content, 482
 - deciding what to do with results, 485
 - external link attraction analysis, 483
 - including in keyword research, 137
 - internal link structure and site architecture, 483
 - link analysis tools, 485
 - of SEO strategy, 484
- competitive intelligence reports from Trellian, 158
- competitive intelligence tools, 121
- comScore
 - Marketer, 163
 - search query length, 3
- Conductor, 486
- conferences on SEO, 505
- consultants in SEO, 515
 - selecting, 521
- contacting potential linkers, methods for, 312
- contains: operator (Bing), 63
- content
 - analyzing for competitors, 482
 - attracting links with, 296
 - CMS and automatically generated content, 219
 - CMS issues, 263
 - contributions by customers or website visitors, 132
 - controlling with cookies and session IDs, 234–238
 - creating link bait or viral content, 317
 - delivery of, and search spider control, 238–254
 - cloaking and segmenting content delivery, 238
 - displaying different content to search engines and visitors, 241
 - when to show different content to visitors and search engines, 240
 - duplicate, 461
 - duplicate content checks, 107
 - duplicate content issues, 226–234
 - avoiding by using cookies and session IDs, 238
 - avoiding duplicate content, 232

- consequences of duplicate content, 227
- copyright infringement, 231
- search engine identification of duplicates, 228
- evaluating resources in planning of SEO, 100
- evaluation by search engines, 32
- hidden, 449–455
 - construed as spam, 454
- importance to SEO goals and objectives, 86
- indexable, 182
- keyword cannibalization and, 112
- marketing for link acquisition, 297
- more searchable content and content types, 538–541
- moving, 439–444
 - mapping moves, 441–443
- optimizing for search engines, 222–226
 - content structure, 222
 - CSS and semantic markup, 223
 - uniqueness and depth of content, 224
- publishing your previously unpublished content, 131
- reviewing for link value, 311
- reviewing in SEO audit, 108
- similar or duplicate of other web pages, 51
- technology decisions affecting SEO, 102
- theft of, 467–470
- user-generated, leveraging in link building, 316
- content syndication, 302, 314
 - duplicate content and, 228
- context and relevance for links, 279
 - evaluating, 292
- conversion optimization, 170
- conversion rates
 - determining for keywords, 167
 - keyboard cannibalization and, 112
- conversions
 - common types of, 388
 - increasing, 394
 - segmenting campaigns and SEO by conversion rate, 393
 - setting up analytics software to track, 390
- cookies, 234–238
 - defined, 234
 - interpretation by search engines, 236
 - search engine problems with, 201
 - turning off and deletion by users, 235
 - using to control search engine access, 237
- copyright
 - for images, 349
 - infringement of, identifying and addressing, 231
 - locating and dealing with infringers, 467–470
 - violations of, 462
- CopyScape.com, 231
- Copyscape.com, 467
- Core Metrics, 377
- cost-per-click (CPC), 150
- countries, targeting a specific country, 275
- country-specific search engines, 74–78
 - optimizing for, 75
 - profile of Internet usage in China, 76
- Covario, 486
- CPC (cost-per-click), 150
- Craigslist, SEO-friendly architecture, 189
- crawl delay, 245
- crawl errors, 416–419
 - search engine tools for checking, 416
 - third-party tools for checking, 417
- crawl issues report, Bing Webmaster Tools, 493
- crawl rate of a site, 125
- crawlers, 30
- crawling and indexing, 30
- crawling improvements by search engines, 538
- Creative Commons (CC) licensed content, 65
- cross-links
 - PageRank calculation, 282
 - using to avoid keyword cannibalization, 219
 - using to distribute link juice, 114
- CSS
 - benefits for SEO and semantic markup, 223
 - exceptions, 265
 - formatting styles, showing different content to search engines and visitors, 241
- CSS image replacement, 182
- CTR (click-through rate), 150
- customer discounts and incentives, 319
- customers and website visitors, content contributions by (see UGC)
- Czech Republic, Seznam search engine, 75

D

- dashboards, web analytics, 383
- data aggregators, 339
- date search operators (Google), 141
- deep links, 110
- deep versus flat site architecture, 194
- Delicious, 323
- Didit, 14
- Digg
 - branding considerations and, 100
 - using in link building campaign, 323
- direct marketing, using SEO for, 93
- directories, 298–301
 - classifying for link value, 300
 - detecting directories that pass link juice, 301
 - essential factors search engines look for, 300
 - list of quality article directories, 314
 - vertical, 340

- disambiguation of search queries, 46
 diversity of search results, 47
 DMCA (Digital Millennium Copyright Act), 467–
 470
 designated agents, 468
 infringement notification, 468
 infringement requests, 231
 doc type search, advanced, 64
 document analysis, 42
 document text, keyword placement in, 217
 Dogpile meta search engine, 72
 domains
 changes in, maintaining search engine visibility,
 445
 international targeting and, 276
 links from, 290
 multiple domains you control, leveraging for
 SEO, 131
 optimizing for SEO, 207
 recognizing low-quality domains, 459
 root domains, subdomains, and microsites, 201–
 207
 using TLD other than .com, 206
 duplicate content issues, 226–234
 consequences of duplicate content, 227
 copyright infringement, 231
 definitions of terms, 226
 duplicate titles and meta tags on many pages,
 52
 filtering and penalties by search machines, 461
 how to avoid duplicate content, 232
 pagination and, 195
 search engine identification of duplicate
 content, 228
 dynamic URLs, 102
- E**
- e-commerce
 action tracking on sites, 384
 conducted without SSL certificate, 461
 SEO for sales, 91
 SSL and HTTPS pages, 233
 using SEO for lead generation and direct
 marketing, 93
- eBay, 341
 email pitch used for direct link requests, 304
 embed tag (HTML), 41
 Enquiero, 14
 Enquisite
 Campaign, 168
 Optimizer, 164, 407
 rankings data, 414
 search analytics, 118
 Exalead search engine, link data, 409
 ext: operator, 55
- external links
 attracting with quality and volume of content,
 100
 attraction of, analysis for competitors, 483
 auditing for SEO problems, 109
 keyword cannibalization and, 112
 to low-quality or spam sites, 52
 search engine tools for reporting, 401
 eye tracking results (user scans of results pages),
 14
- Eyetools, 14
- F**
- Facebook, 324
 SEO practitioners on, 504
 fans of your business, contributions from, 132
 feed: operator (Bing), 63
 feedmaps, 346
 filenames for images, 217
 filetype: operator, 55, 64
 Firefox
 extensions for Yahoo! link: and linkdomain:
 operators, 61
 plug-ins accessing Google advanced search
 queries, 59
 SearchStatus plug-in, 402
 viewing source code, 35
 First Click Free concept, 240, 539
 first stages of SEO (see SEO projects, first stages)
 Flash
 links in, text-link-based alternative to, 200
 links or content based in, 102
 optimizing for SEO, 269–274
 Adobe Flash search engine SDK, 270
 disclaimer for SWFObject and NoScript
 methods, 272
 internal Flash coding, 270
 meta tags, 270
 NoScript tags, 274
 scalable Inman Flash Replacement (sIFR),
 274
 SWFObject tags, 273
 search engine reading of Flash files, 40
 flat versus deep site architecture, 194
 Flickr, 328
 image search optimization through,
 350
 videos on, 369
 fonts, choice of, sizing, and contrast issues, 222
 Foresee/ACSI Report for eMarketer (August 2007),
 9
 formating of content, 222
 forms
 content behind, 102
 forcing form submission to block search engine
 robots, 253

- links in submission-required forms, 183
 - search and, 196
 - forums
 - action tracking on, 386
 - automated link dropping on, 308
 - exploitation to gain links, 293
 - interacting with search engine representatives on, 475
 - manual creation of links on, 305
 - NoFollow on links, 309
 - participation in, 324
 - SEO, 502
 - using to attract links, 316
 - frames, 41
 - links in, 183
 - problems for search engines, 199
 - frequency of recent keyword usage, 140
 - freshness in search results, 48
 - fully qualified domains (FQDs), 202
 - future of SEO, 531–556
 - increased market saturation and competition, 552
 - increasing importance of local, mobile, and voice recognition search, 546–552
 - major objectives of SEO, 554
 - more searchable content and content types, 538–541
 - ongoing evolution of search, 532–537
 - fuzzy logic, 43
 - Fuzzy Set Theory, 43
- G**
- gender differences in searching, 87
 - geolocation SEO audit, 109
 - GetListed, 344
 - giveaways as link incentives, 302
 - goals and objectives for SEO, 79–96
 - business factors affecting SEO plan, 84
 - considerations using SEO for raw traffic, 91
 - customizing your SEO plan, 82
 - developing SEO plan prior to site, 84
 - e-commerce sales, 91
 - ideological influence, 95
 - lead generation and direct marketing, 93
 - mindshare/branding, 92
 - reputation management, 93
 - strategic goals SEO can fulfill, 80
 - understanding search engine traffic and visitor intent, 83
 - understanding your audience and finding your niche, 85–90
 - Google
 - advanced search operators, 54
 - combined, 57
 - Firefox plug-ins for, 59
 - AdWords Keyword Tool, 142
 - Analytics, 117, 377
 - conversion tracking, 390
 - traffic reports from, 127
 - Blog Search, 409
 - blog search counts, 138
 - Blogoscopy, 423
 - Chrome, 543
 - cloaking, position on, 238
 - cloud computing, 545
 - comparison of search result with Yahoo!, 481
 - DMCA notices to, 468
 - dominance in web search, 536
 - First Click Free program, 240
 - growth of search complexity, 532
 - guidelines for recognizing search engine spam, 461
 - guidelines on search results in the search results, 225
 - historical information application, 430
 - identifying duplicate content, 228
 - Local Business Center, 340
 - market share, 2, 75
 - Mobile and Googlebot-Mobile, 367
 - most popular domains in February 2009, 336
 - most popular domains in May 2006, 334
 - News, 363
 - submission details, 365
 - paid links, reporting to, 320
 - penalties, identifying, 464
 - policy on paid links, 306
 - Product Search, 352
 - getting into, 353
 - performance reporting, 354
 - punishing sites that sell or buy links, 321
 - rank checking, problems with, 126
 - Reader, blog subscription data, 423
 - reinclusion requests, 466
 - sandbox, 53
 - search results layout, 24
 - SearchWiki, 544
 - site indexing data, 398
 - Traffic Estimator, 144
 - Translate, 65
 - Trends
 - keyword research, 160
 - seasonal data on keyword demand, 176
 - Trends for Websites, 121, 428
 - Universal Search, 15, 335
 - updating Google Sitemap, 186
 - vertical results in SERPs, 26
 - vertical search, 66
 - Blog Search, 69
 - Book Search, 70
 - Maps, 68

- specialty search engines, 73
 - Universal Search, 70
 - Webmaster Central Blog, 474
 - Webmaster Guidelines, 456
 - for hidden text, 454
 - Webmaster Tools, 117, 487–492
 - checking backlinks, 399
 - crawl data from, 125
 - crawl errors, 416
 - rankings data, 414
 - site indexing data, 398
 - spider activity report, 426
 - Google Bombing, 285
- H**
- hasfeed: operator (Bing), 63
 - heading tags, 216
 - health user search session (example), 11
 - hiring SEO talent, 517–519
 - example job opening, 518
 - pitching the person, 517
 - selecting the right practitioner, 517
 - Hitwise, 121
 - competitive analysis, 486
 - news search market share, 363
 - Search Intelligence keyword research tool, 161
 - search query length data, 4
 - host domain, 202
 - hosting issues
 - CMS third-party add-ons and, 268
 - negatively impacting search engine rankings, 115
 - potential problems with host IP address, 201
 - HTML
 - embed tag, 41
 - source code for web pages, 34
 - tag customization for SEO, 264
 - HTTP header checks, 108
 - HTTP status codes for redirects, 255
 - HTTPS pages, 233
 - canonicalization issues with rewrites of URLs, 260
 - robots.txt file, 244
 - hubs and authorities, 287
- I**
- Ice.com, YouTube campaign, 326
 - ideological influence, SEO for, 95
 - iframes, 41
 - links in, 183
 - problems for search engines, 199
 - using to prevent indexing of content, 252
 - IIS servers
 - ISAPI_Rewrite, 440
 - redirecting URLs, 256
 - rewriting URLs with ISAPI_Rewrite, 257–261
 - image search, 66
 - optimizing for, 346–352
 - optimization tips, 348
 - through Flickr and other image sharing sites, 350
 - image sharing sites, 350
 - images
 - alt attribute (HTML), 38
 - filenames and alt tags, 217
 - making accessible to crawlers, 198
 - making visible to search engines, 182
 - SEO issues, 265
 - importance rankings, 32
 - in-house SEO, 507
 - mixing with outsourced, 528
 - in-house SEO specialist, 510
 - inanchor: operator, 55, 64
 - inbody: operator (Bing), 63
 - inbound links
 - anchor text, 50
 - Bing Webmaster Tools report on, 495
 - on blogs, 362
 - topical relevance of, 51
 - incentive-based link marketing, 319
 - incentive-based link requests, 302
 - income (household), Internet use by, 4
 - indexes
 - inclusion in search engine indexes, 289
 - index-to-crawl ratio, 437
 - redirecting index file without looping, 261
 - removing URLs from search engine index, 254
 - search engine, keeping pieces of page content out of, 237
 - indexing, 30
 - benchmarking current status, 124
 - index component of meta robots tag, 249
 - indexable content on sites, 182
 - multimedia becoming indexable, 540
 - Noindex directive, robots.txt, 246
 - preventing by use of iframes, 252
 - site indexing data for search engines, 397
 - influencers, identifying, 421
 - info guides, local, 341
 - info: operator, 56
 - information architect, 187
 - information architecture
 - optimizing for search engine crawlers, 187
 - deep versus flat architecture, 194
 - design principles, 191
 - logical, category-based flow, 187
 - navigation, 196
 - information retrieval (IR), 41
 - common types of searches, 43

- informational searches, 6
 - InfoUSA, 339, 546
 - instant answers, 26–29
 - intent of searchers, discovering, 5–9
 - intent of searchers, understanding, 83, 542
 - interactive content, 541
 - internal links, 51, 115
 - (see also links)
 - anchor text on, customizing, 264
 - auditing for SEO problems, 108
 - fixing problem with, 113
 - keyword-rich anchor text, 105
 - NoFollow attribute, 247
 - international targeting
 - basics of, 275
 - best practices, 275
 - Internet Explorer, viewing source code, 35
 - Internet usage in China, 76
 - Internet usage statistics, 552
 - Internet users by household income, 4
 - intext: operator, 55
 - intitle: operator, 55, 64
 - searching for duplicate content, 107
 - Intuit, YouTube campaign, 326
 - inurl: operator, 54
 - searching for duplicate content, 107
 - IP addresses
 - blocked, negative impact on SEO, 116
 - blocking bots and cloaking by, 252
 - problems with IP address of host, 201
 - shared, negative impact on SEO, 115
 - ip: operator (Bing), 63
 - iPhone
 - Google Universal Search results for, 335
 - mobile search results, 366
- J**
- Java, search engine crawlers and, 197
 - JavaScript, 275
 - browsers without, Flash .swf files and, 274
 - hosted solutions and CMS third-party add-ons, 268
 - links or content based in, 102
 - noscript HTML tags and, 39
 - problems for search engine crawlers, 198
 - use by web analytics platforms, 377
 - using to track product sales, 391
 - JetBlue Airlines, use of YouTube, 326
 - job searches, 70
- K**
- keyword cannibalization, 111, 218
 - keyword mapping, 111
 - keyword research, 135–179
 - completing for site architecture audit, 110
 - data from search engines, 138
 - determining keyword value using landing page optimization, 170
 - domain expertise and site content analysis, 136
 - estimating value, relevance, and conversion rates, 166
 - leveraging long tail of keyword demand, 171–176
 - long tail of search demand curve, 136
 - testing ad campaigns and third-party search data, 168
 - tools, 142–166
 - comScore Marketer, 163
 - considerations with, 166
 - Enquisite Optimizer, 164
 - Google AdWords Keyword Tool, 142
 - Google Traffic Estimator, 144
 - Google Trends, 160
 - Hitwise, 161
 - KeywordDiscovery, 154–160
 - Microsoft adCenter Keyword Generation Tool, 149
 - Wordtracker, 151
 - Yahoo! Search Marketing, 148
 - trending, seasonality, and seasonal fluctuations in demand, 176
 - keyword reviews, 111
 - (see also keyword research; keywords)
 - auditing existing site for SEO problems, 110
 - keyword stuffing, 50, 211
 - keyword targeting, 93, 211–222
 - avoiding keyword cannibalization, 218
 - boldface text, 218
 - in CMSs and automatically generated content, 219
 - by content creators, 220
 - document text, 217
 - heading tags, 216
 - image filenames and alt tags, 217
 - long tail of keyword demand, 220
 - meta description tags, 214
 - title tags, 212
 - KeywordDiscovery, 154–160
 - free search term suggestion tool, 159
 - keyword data by search engine, 177
 - seasonal data on keyword demand, 177
 - keywords
 - brand-to-non-brand ratio, 435
 - competitiveness of market, 64
 - health check, 107
 - impact of using in blog page titles, 358
 - matching up to site hierarchy, 103

- page performance by, Bing Webmaster Tools, 495
- ratio of keywords to pages yielding search traffic, 436
- search operators with, 53
- search traffic by keyword, 379
- search visitors per keyword, 436
- tracking conversions by, 393
- use in body text, 51
- use in title tag, 50
- L**
- landing page optimization, 170
- languages
- Google Trends information on searches in, 161
 - issues with multiple languages, 277
 - selecting for Traffic Estimator, 147
- latent semantic analysis (LSA), 44
- latent semantic indexing (LSI), 44
- layout of content, 223
- lead generation, using SEO for, 93
- length of content, 223
- licensed content, duplicate content issues, 230
- Lifeinsure.com, link bait content, 318
- limited cash options for SEO, 514
- link analysis, 45
- external link attraction for competitors, 483
 - search engine judging of content and, 34
- link bait bumps, 394
- link baiting, 296
- creating link bait content, 317
- link brokers, 293, 306
- link building, 296–318
- buying links, 306
 - choosing right strategy, 309–313
 - content syndication, 314
 - creating link bait or viral content, 317
 - direct link requests, 303
 - incentive-based link requests, 302
 - keys to successful link building, 295
 - leveraging user-generated content, 316
 - manual links, creating with social media, 304
 - marketing content for, 297
 - measurements for campaigns, 399
 - NoFollow uses and scams, 309
 - social networking for, 321–330
 - summary of process, 312
 - third-party tools for, 407
 - using automated link dropping, 308
 - using content to attract links, 296
 - using directories, 298–301
 - using link farms or link networks, 307
 - using spammy giveaways, 308
- Link Diagnosis, 310, 485
- information about competitors' links, 404
- link farms, 307
- link juice distribution, fixing problem with, 113
- link marketing, incentive-based, 319
- link neighborhoods, 45
- identifying, 65
- link networks, 308
- link schemes, 52
- link spam, search engine tactics against, 319
- link: operator (Yahoo!), 60
- Firefox extension for, 61
- linkdomain: operator (Yahoo!), 60
- Firefox extension for, 61
- LinkedIn, 304, 324
- SEO practitioners on, 504
- linkfromdomain: operator (Bing), 63
- LinkMaps, 407
- links, 279–331
- absolute links, 233
 - anchor text in internal links, 105
 - based in JavaScript, Java, or Flash, 102
 - to blog posts, 423
 - blog, NoFollow attribute on, 360
 - categorizing value of, 303
 - competitive link analysis tools, 485
 - crawlable link structure for search engines, 182
 - cross-linking relevant content, 103, 219
 - determining value of, 294
 - external linking SEO audit, 109
 - finding sites engaging in link schemes, 118
 - identifying sources of high-quality links, 98
 - important ranking factors, 50
 - influence on search engine rankings, 280–289
 - anchor text, 285
 - authority, 286
 - how search engines use links, 288
 - original PageRank algorithm, 280–284
 - relevance, 286
 - trust, 287
- internal link structure, competitive analysis of, 483
- internal linking checks in SEO audit, 108
- keys to getting links, 295
- making your site attractive to link to, 188
- measuring value of, 409
- minimizing link depth on a site, 105
- multiple navigation paths and sculpting flow of
- link juice, 237
- negative ranking factors, 51
- NoFollow attribute, 246
- number of links per page crawled by search engines, 195
- obtaining and managing for blogs, 360
- problems posed by microsites, 205
- rate of acquisition, 52

- reasons for creating, 295
 - search engine evaluation of, additional factors, 289–294
 - context and relevance, 292
 - linking domains, 290
 - source diversity, 291
 - source independence, 289
 - temporal factors, 291
 - supporting local search, 346
 - temporal link growth measurements, 430
 - text-link-based navigational structure, 200
 - third-party measuring tools, 403–409
 - tools for checking backlinks, 399
 - web crawling and indexing, 30
 - Linkscape, 310, 404, 485
 - Live HTTP Headers, 108
 - load time for pages, 110
 - local business profiles, 342
 - management services for your data, 344
 - local search, 14
 - Google Maps, 68
 - optimizing for, 338–346
 - checking your local listings, 339–342
 - local business profiles, 342
 - management services for local profile information, 343
 - optimizing your website, 344
 - website optimization, 344
 - server geography and, 116
 - Local Search Guide, 340
 - Local.com, 546
 - Localeze, 344, 546
 - location, 68
 - (see also local search)
 - geographic location of searcher, 84
 - geolocation lookup by search engines, 53
 - geolocation SEO audit, 109
 - impact on searches, 88
 - location: operator (Bing), 63
 - logfile analyzers, 117, 425
 - login
 - blocking search engine robots, 254
 - content behind, exposing to Google spiders, 240
 - content requiring, 102
 - keeping page content out of engines' indexes, 237
 - links in submission-required forms, 183
 - pages requiring, granting access to, 237
 - long tail of keyword demand, 171–176
 - editorial content strategies targeting, 175
 - extracting terms from relevant web pages, 171
 - identifying patterns in, 172
 - mining keyword research tools for data on, 172
 - targeting, 220
 - in traffic by keyword reports, 380
 - UGC-generated content strategies for targeting, 176
 - long tail of search, 136
 - long tail SEO, key performance indicators, 435
 - low-cost tools, leveraging for SEO, 513
 - low-quality sites, external links to, 52
 - LSA (latent semantic analysis), 44
 - LSI (latent semantic indexing), 44
 - Lyris HQ Web Analytics, 425
- M**
- magazines covering SEO issues, 474
 - Majestic-SEO, 310, 485
 - map searches, 68
 - map spam, 548
 - MapQuest, 546
 - market
 - determining degree of competitiveness in, 98
 - increased saturation and competition in search market, 552
 - mobile search, 548
 - Marketer tool (comScore), 163
 - marketing
 - incentive-based link marketing, 319
 - link building as, 312
 - SEO for lead generation and direct marketing, 93
 - measurements for SEO projects, 116, 373–437
 - assessing historical progress, 122
 - competitive and diagnostic search metrics, 397–434
 - Google and Bing Webmaster Tools, 117
 - logfile tracking, 117
 - measuring search traffic, 376–388
 - other useful metrics, 477
 - search analytics, 118
 - testing SEO, 475
 - tying SEO to conversion and ROI, 388–397
 - attribution, 389
 - Web Analytics, 116
 - why measuring success is essential, 374–376
 - Meetup.com, 329
 - Merrell shoes user search (example), 11
 - meta description tags, 38
 - meta robots tag, 248
 - meta search engines, 72
 - meta tags, 36
 - customization for CMS content, 264
 - description
 - guidelines for, 214
 - duplication of, negative ranking factor, 52
 - reviewing in SEO audit, 108
 - MetaCrawler, 72

Metcalf's Law, 549
 microsites, 204
 reasons for not using, 205
 reasons for using, 204
 Microsoft
 adCenter Keyword Generation Tool, 149
 IIS web servers, URL redirecting, 256
 Word, 513
 mindshare/branding, SEO for, 92
 mobiForge, 549
 mobile search, 548–551
 optimization, 366
 U.S. market, 548
 U.S. mobile search market share, 551
 worldwide mobile Internet growth, 550
 Moveable Type, 357
 MRSS feeds, acceptance by video sites, 369
 MultiChannel Merchant, 394
 multimedia becoming indexable, 540
 MySpace, 324
 videos on, 369

N

natural versus paid search results, click tracking
 users, 16–22
 navigation, 51
 (see also links)
 breadcrumb navigation, 105
 content evaluation and, 33
 search-engine-friendly, guidelines for, 200
 search-friendly, for websites, 196
 showing multiple navigation paths and
 sculpting flow of link juice, 237
 vertical, 24
 navigational searches, 6
 neighborhoods, link, 45, 65
 Netconcepts, 119
 GravityStream, 268
 metrics for long tail search traffic, 435–437
 proxy optimization technology, 512
 SEO Title Tag, 358
 theft of its site design, 467
 NetConcepts
 GravityStream, 450
 network effect, 549
 news search, 68
 optimization for vertical search, 363
 newspapers, 341
 NoFollow attribute, 183, 246
 on blog links, 360
 links in visitor comments in CMS content, 267
 links to social media, 304
 uses and scams, 309
 NoFollow meta tag, 183, 232
 search engine robot's response to, 249

Noindex directive (robots.txt), 246
 NoIndex meta tag
 for low-value pages, 266
 search engine robot's response to, 249
 NoIndex metatag, 232
 noscript tags, 39
 NoScript tags
 addressing problems with JavaScript and AJAX, 275
 number range operator (...), 64

O

Omniture, 377
 onebox, 26, 352
 online classified sites, 341
 ontology of websites, 193
 opportunities, identifying, 133
 OR operator, 54
 originalextension: operator (Yahoo!), 60
 outside expert in SEO, case for working with, 519
 outsourced SEO, 508

mixing with in-house SEO teams, 528
 selecting an SEO firm or consultant, 521–528
 communicating with candidate firms, 526
 getting process started, 521
 making the decision, 527
 preparing a RFP, 522–525
 for small organizations, 511

P

page load time, 110
 Page Strength tool, 411
 PageRank
 Google crawl rate based on, 289
 link value measurements, 412
 original algorithm, 280–284
 using to determine link value, 294
 pagination, 195, 265
 paid links (see buying links)
 paid search, 80
 (see also PPC advertising)
 estimates on competitiveness, 168
 Google keyword research tools, 142–148
 natural results versus, click tracking users, 16–22
 partnerships, leveraging for SEO, 131
 password protection, blocking search engine
 robots, 254
 pattern matching
 rewriting URLs, 258–261
 in robots.txt file, 245
 pay per click (see PPC advertising)
 pay-level domains (PLDs), 202
 penalties levied by search engines, 2, 120, 456

(see also spam)
 for copyright infringement, 231
 for duplicate content, 227
 identifying, 464
 for keyword stuffing, 211
 reinclusion or reconsideration requests, 464
 spam filtering and, 455
 performance reporting on Google Product Search, 354
 personalization of search, 53, 541
 photo sites, stock, 349
 phrases common in searches, 140
 pinging RSS feed services, 265
 Pingomatic.com, 357
 planning for SEO, 97–100
 branding considerations, 100
 content resources, 100
 identifying sources of high-quality links, 98
 market segmentation, 98
 technology choices, 98
 understanding your competition, 100
 PLDs (pay-level domains), 202
 popularity rankings, 32
 PPC (pay per click) advertising, 9
 click tracking, natural search results versus, 16–22
 determining CTR and CPC, 149
 paid links and, 306
 SEO versus, 80
 spending on, SEO versus, 20
 presell pages, 293
 print pages, 233
 product search, optimizing for, 352–354
 getting into Google Product Search, 353
 products and services, mapping, 85
 Propeller, 323
 proximity searches, 43
 public relations campaigns, tracking, 421

Q

Quantcast, 121
 site data, 429
 traffic measurements, 427
 query deserves freshness (QDF), 48

R

R.H. Donnelly, 344
 ranking factors, 32
 important, 49
 negative, 51
 server and hosting issues, 115
 other, 52
 rankings, 31
 analysis of top-ranking sites and pages, 478

benchmarking current rankings, 126
 data on, uses of, 413
 factors specific to mobile SEO, 368
 role of links in, 289
 tools to obtain data on, 414
 Recaptcha.net, 540
 reciprocal links, dangers of, 302
 reconsideration requests, 466
 Reddit, 323
 redirects, 254–262
 301 HTTP status code, 440
 301 redirect versus canonical URL tag attribute, 250
 checking for canonical redirect, 107
 checking in SEO audit, 108
 creating spider traps, 200
 for domain name change, 446
 good and bad implementations of, 255
 redirecting index file without looping, 261
 SEO-friendly websites, 190
 temporary (302) redirects, 102
 URL redirecting, 256
 referral source, action tracking by, 394
 referring sites report, 381
 refinements for search queries, 26
 region: operator (Yahoo!), 60
 regular expressions (and pattern matching)
 common errors in, 258
 greedy matching, avoiding, 259
 rewriting URLs, 258
 special characters, 258
 reinclusion requests, 465
 rel="NoFollow" attribute (see NoFollow attribute)
 related terms, 139
 related: operator, 55, 65
 relationship building, offline, 325
 relevance, 32
 identifying relevant keywords, 167
 of links, 286, 292
 REP (Robots Exclusion Protocol), 416
 reputation management
 responding to user ratings, 345
 SEO for, 93
 reputation monitoring, 421
 research and study in SEO, 473–506
 analysis of algorithmic differentiation for engines and search types, 480
 analysis of top-ranking sites and pages, 478
 competitive analysis, 482
 participation in conferences and organizations, 505
 resources for SEO, 473–475
 SEO industry on the Web, 500–504
 testing SEO, 475–478
 using search engine SEO tools, 487–500

- rewriting URLs, 256
 - using mod_rewrite for Apache and ISAPI_Rewrite for IIS, 257–261
 - RFP (request for proposal), preparing, 522–525
 - robots
 - pages blocked by REP, 416
 - search engine robot traffic analysis, 425–426
 - robots tag, 248
 - robots-nocontent tag (Yahoo!), 237
 - robots.txt file
 - syntax, 243–246
 - using to control spider access on your site, 241
 - using to prevent crawling of duplicate pages, 232
 - verification of, 108
 - ROI (return on investment)
 - determining for SEO projects, 396
 - increasing using SEO, 81
 - root domain, 202
 - separate, when to use, 204
 - RSS feeds
 - blog users subscribing to, 421
 - optimization for vertical search, 355
 - optimizing for vertical search
 - other considerations, 356
 - personalized, 357
 - submission to Yahoo! News, 366
 - tracking and measurement, 356
 - video sites that accept, 369
 - Russia, Yandex search engine, 75
- S**
- Safe Harbor provision, 469
 - scalable Inman Flash Replacement (sIFR), 274
 - search analytics, 118
 - Search Engine Land, 474
 - Stats & Behaviors page, 5
 - search engine optimization (see SEO)
 - search engine rankings (see ranking factors; rankings)
 - Search Engine Roundtable, 316
 - Search Engine Watch, 474
 - data on cross-channel interactions, 389
 - search engines, 23
 - (see also Bing; Google; Yahoo!; SERP)
 - advanced search techniques, 53
 - algorithm-based ranking systems, 30–41
 - content invisible to search engines, 40
 - content visible to search engines, 34
 - crawling and indexing, 30
 - retrieval and rankings, 31
 - web page content evaluation, 32
 - algorithmic differentiation across, analysis of, 480
 - analysis of ranking factors, 49–53
 - breakout of traffic by, 379
 - canonical tags, 250
 - commentary from employees, 474
 - country-specific, 74–78
 - crawling improvements, 538
 - differences among, understanding, 482
 - document analysis and semantic connectivity, 42
 - health check in site audit, 106
 - how they drive Web commerce, 14
 - indexing data from, 124
 - keyword research tools, 138
 - link analysis, 45
 - maintaining visibility through domain name changes, 445
 - maintaining visibility through site redesigns, 444
 - market share, 2
 - mission of, 2
 - ongoing evolution of search, 532–537
 - problem words, disambiguation, and diversity, 46
 - reasons for failure of search algorithms, 49
 - robot traffic analysis, 425–426
 - search becoming personalized and user-influenced, 541–546
 - trust measurement for sites, 288
 - understanding results, 24–29
 - vertical search, 66–73
 - search functionality on websites, 196
 - Search Intelligence tool (Hitwise), 161
 - Search Marketing keyword tool (Yahoo!), 148
 - search operators, 53–65
 - Bing, 63
 - Google, 54–59
 - Google date search operators, 141
 - more advanced techniques, 64–65
 - Yahoo!, 59–63
 - ~ (tilde), 44
 - search patterns, new, 543
 - search phrases, information on, 381
 - search results, site pages perceived as, 225
 - search traffic, measuring, 374, 376–388
 - using web analytics, 377
 - action tracking, 384–387
 - analytics dashboards, 383
 - not getting lost in the data, 387
 - segmenting traffic with multiple parameters, 381
 - traffic by search engine, 379
 - searcher intent, determining, 5–9, 83, 542
 - searches
 - common types in IR, 43
 - example search queries and intent, 8
 - informational queries, 6

- by market segment, 9
- navigational queries, 6
- query length, 3
- transactional queries, 7
- searching
 - eye tracking, user scanning of results pages, 14
 - how people search, 9
 - human goals of, 3
 - searchers and information sought, 4
- SearchStatus plug-in (Firefox), 59, 402
- seasonality in keyword demand, 160, 176
- selling links, 52, 303
 - (see also buying links)
- semantic connectivity, 43
- semantic maps, 33
- semantic markup, 224
- SEO (search engine optimization), xxi
 - in-house versus outsourced, 507–509
 - solutions for large organizations, 514–516
 - solutions for small organizations, 509–511
 - working with limited resources or budget, 511–514
- SEO Browser, 110
- SEO culture, building into your organization, 529
- SEO goals and objectives (see goals and objectives for SEO)
- SEO industry on the Web, 500–504
 - blogs, 500–502
 - communities in social networks, 504
 - forums, 502
- SEO problems, auditing site for (see audits)
- SEO projects, first stages, 97–134
 - auditing existing site for SEO problems, 106–116
 - benchmarking current indexing status, 124
 - benchmarking current rankings, 126
 - benchmarking current traffic sources and volume, 127–130
 - conducting SWOT analysis, 132
 - defining site information architecture, 101–106
 - determining top competitors, 118–122
 - identifying server statistics software and gaining access, 116
 - identifying site development process and players, 101
 - leveraging business assets for SEO, 131–132
 - measuring results from SEO efforts, 122–124
 - planning, 97–100
- SEO research and study (see research and study in SEO)
- SEO-friendly websites, 181–278
 - CMS issues, 263–268
 - content optimization, 222–226
 - duplicate content issues, 226
- keyword targeting, 211–222
- making site accessible to search engines, 181–187
- optimal information architecture, 187–201
- optimizing domain names and URLs, 207–211
- optimizing Flash, JavaScript, and AJAX, 268–275
- redirects, 254–262
 - good and bad implementations, 255
 - index file redirects without looping, 261
 - methods of URL redirecting and rewriting, 256–261
- root domains, subdomains, and microsites, 201–207
- targeting multiple languages and countries, 275
- SEOMoz, 474
- action tracking in sign up form, 384
- article on failures of search algorithms, 49
- Crawl Test Tool, 417
- Keyword Difficulty Tool, 168
- Linkscape, mozTrust, 288
- Page Strength, 411
- ranking factors survey, 49
- search engine ranking factors study, 212
- SERP (search engine results pages), 24
 - click tracking, natural versus paid SERP
 - distribution of results and traffic, 17
 - intents and, 18
 - interaction of natural and paid search, 20
 - other factors, 20
 - click tracking, natural versus paid SERPs, 16
 - layout of, 24
 - vertical results in, 26
- SERPs (search engine results pages)
 - importance of, 16
- servers
 - changing, 447
 - monitoring after server move, 448
 - chaning
 - scenarios similar to server moves, 448
 - issues negatively impacting search engine rankings, 115
- session IDs, 234–238
 - defined, 234
 - interpretation by search engines, 236
 - search engine problems with, 201
 - using to control search engine access, 237
- shingles, 226
- sIFR (scalable Inman Flash Replacement), 274
- site architecture (see information architecture)
- site: operator, 54
- Sitemaps (see XML Sitemaps)
- SMART objectives, 89
- smartphones, U.S. adoption of, 549

- Smirnoff's Raw Tea campaign on YouTube, 327
 snippets, 226
 social media
 manual linking to, 304
 using to contact potential linkers, 312
 using to market content, 297
 social networking for links, 321–330
 blogging, 322
 social news and tagging sites, 323
 offline relationship building, 325
 participation in forums and social network, 324
 success stories using YouTube, 325
 tips for more sites, 327
 social networks, SEO communities in, 504
 sort orders, multiple, 233
 source diversity of links, 291
 source independence of links, 289
 South Korea, Naver search engine, 75
 spam, 2
 anchor text in internal links, 109
 automated link dropping, 308
 email, unsolicited, 304
 examples of spam sites and link schemes, 118
 external links to spam sites, 52
 filtering and penalties by search engines, 455–467
 competitors reporting you, 461
 duplicate content, 461
 identifying penalties, 464
 reinclusion/reconsideration requests, 464
 rules for spam-free SEO, 462
 signals of low-quality domains and spam sites, 458–461
 filtering spam domains with Google sandbox, 53
 hidden content construed as, 454
 how search engines fight link spam, 319
 issues with pingbacks, trackbacks, and comments, 267
 map spam in local search results, 548
 search engines' battle against, 533
 spammy giveaways, 308
 Sphinn, 504
 spider traps, 200
 spiders, 30
 data on activities, 425
 learning semantic relationships, 192
 site accessibility to, 106
 site content problematic for, 196
 Squid, 408
 SSL
 HTTPS pages and, 233
 search engine robot crawling of HTTP and HTTPS pages, 244
 staff member changes for SEO team, 470
 sticky posts, 359
 stock photo sites, images from, 349
 Stone Temple Consulting, 33
 interview series, 380
 long tail keywords report, 380
 meta tags in HTML source, 37
 Web Analytics Shootout report, 378
 strengths, identifying, 132
 structural decisions for websites, 103–106
 cross-linking relevant content, 103
 site architecture, 110
 targeting keywords, 103
 using anchor text in internal links, 105
 using breadcrumb navigation, 105
 StumbleUpon, 323
 subdomains, 202
 avoiding unnecessary, 109
 search engine accessibility and, 190
 when to use, 203
 subfolders, 202
 when to use, 202
 subscriber-only or deleted content, uncovering, 65
 success, measuring for SEO efforts, 373
 Superpages, 546
 SWOT (strengths, weaknesses, opportunities, threats) analysis, 88
 conducting SEO and website analysis, 132
 syndication of content, 297

T

- tag clouds on blogs, 361
 tagging sites, 323
 tags, use for marketing on YouTube, 327
 taxonomy and ontology of websites, 193
 technology choices, SEO and, 98
 important issues to address, 102
 Technorati, blog reactions data, 409, 423
 temporal factors in link evaluation, 291
 Teoma algorithm, 286
 term weighting, 43
 testing SEO, 475–478
 other useful SEO metrics, 477
 sample experimental approach, 476
 starting with a hypothesis, 478
 text
 boldface, 218
 code to text ratio, 223
 creating hidden text, 454
 document text and keyword targeting, 217
 unintentionally creating hidden text, 455
 thin affiliate sites, 232, 462
 thin content, 224
 threats, identifying, 133
 three-way link swaps, 308

- Thumbnails, rankings data, 414
- t**
- title tags
- auditing for SEO problems, 108
 - best practices for constructing, 212
 - customization and rules, 264
 - duplicate, negative ranking factor, 52
 - keyword cannibalization and, 218
 - keyword use in, 50
 - optimizing for video search, 371
 - optimizing in blogs, 358
- TLD (top-level domain), 202
- other than .com, when to use, 206
 - source TLDs in link evaluation, 293
 - spam signals and, 461
- tracking SEO results, 373
- (see also measurements for SEO projects)
 - tracking cycle, 375
- traffic, 376
- (see also search traffic, measuring; website traffic)
 - drivers of SEO results, 374
 - search results distribution, paid and natural, 17
 - understanding search engine traffic and visitor intent, 83
- Traffic Estimator (Google), 144
- transactional searches, 7
- Translate (Google), 65
- travel user search session (example), 12
- Trellian
- competitive intelligence reports, 158
 - keyword data, 159
- trends in keyword demand, 176
- triangular link swapping, 308
- trust
- links and, 287
 - role in search algorithms, 31
- TrustRank, 288
- Truveo, 369
- Twitter, 324
- SEO practitioners on, 504
 - using to promote your business, 329
- TypePad, 357
- U**
- UGC (user-generated content), 132
- duplicate content issues, 233
 - leveraging in link building, 316
 - sites based on, action tracking on, 386
 - strategies for long tail targeting, 176
- Unica Affinium NetInsight, 377
- robot report, 425
- unique content, 226
- Universal Search, 70, 335
- url: operator (Yahoo!), 61
- URLs
- canonical URL tag attribute, 250
- checking for clean, short, descriptive URLs, 107
- content moves, 439–444
- factors affecting SEO, 102
- optimizing for search engine crawlers, 190
- redirecting, methods for, 256–261
- removing from search engine index, 254
- selecting optimal URLs for site pages, 209
- SEO issues in CMS content, 265
- static and keyword-rich, 264
- usability
- auditing site for SEO problems, 106
 - logical, category-based flow, 187
 - site usability and search friendliness, 188
- usage data as ranking factor, 53
- user agents
- blocking and cloaking by, 252
 - search engine robot identification by, 243
 - serving different content based on, 238
- user data as ranking factor, 53
- user intent (see searcher intent, determining)
- user interactions with search engines, 543
- user ratings, local search rankings and, 345
- user-driven search results, 544
- V**
- value of keywords, determining, 166
- value proposition for direct link requests, 303
- vendor changes for SEO, 470
- vertical results, 26–29
- vertical search, 66–73, 333–372
- from major search engines, 66–70
 - mobile search optimization, 366
 - opportunities in, 333–338
 - optimizing for image search, 346–352
 - image optimization tips, 348
 - optimizing for local search, 338–346
 - optimizing for news, blog, and feed search, 355–366
 - blog optimization, 357–363
 - news search, 363–366
 - RSS feed optimization, 355–357
 - RSS feed tracking and measurement, 356
- optimizing for product search, 352
- from specialized search engines, 72
- Universal Search/Blended Search, 70
- video search optimization, 368–372
- video search engines
- that crawl your site, 369
 - video upload sites, 369
- video search optimization, 368–372
- video searches, 67
- video, problems for search engine crawlers, 197
- virally spreading link bait content, 318

visibility (branding), using SEO, 80
 visual layout of content, 223
 voice recognition search, 551

W

W3C Web Accessibility initiative page, 39
 Wayback Machine, 121
 helping prove original author of content, 470
 weaknesses, identifying, 132
 Web Accessibility Initiative page (W3C), 39
 web analytics, 376–380
 action tracking in, 384–387
 avoiding getting lost in the data, 387
 choosing an analytics package, 377
 common mistakes with, 387
 establishing current traffic sources and volume, 127–130
 free tools, 377
 packages performing log analysis, 425
 referring sites report, 381
 search engine tools for, 116
 segment search traffic with multiple parameters, 381
 setting up software to track conversions, 390
 signals of spam sites and low-quality domains, 460
 traffic by keyword, 379
 using analytics dashboards, 383
 using as business case for SEO, 376
 valuable SEO data in
 traffic by search engine, 379
 web crawling, 30
 Web documents, indexing (see indexes; indexing)
 web page for this book, xxiii
 Webalizer, 377
 logfile analysis, 117
 Webmaster Tools, 117
 WebmasterWorld Forums, 316
 website traffic
 benchmarking current sources and volume, 127–130
 comparing with competitors, 426–430
 increasing with SEO, 81
 SEO for raw traffic, 91
 websites
 accessing previous versions of competitors' home pages, 121
 analysis of top-ranking sites and pages, 478
 auditing existing site for SEO problems, 106
 keyword reviews, 110
 defining information architecture, 101
 structural decisions, 103
 technology decisions affecting SEO, 102
 identifying site development process and players, 101

maintaining timeline of site changes, 122
 optimizing for local search engines, 344
 previous SEO work, 124
 redesign, maintaining search engine visibility through, 444
 SEO resources, 474
 SEO-friendly (see SEO-friendly websites)
 spam sites, signals of, 458–461
 types of changes that can affect SEO, 123
 Webtrends, 377, 425
 widgets
 links embedded in, 302
 spammy giveaways in link building, 308
 Wikipedia, using for link building, 327
 wikis, 328
 WordPress, 357
 plug-ins, 360
 Wordtracker, 151
 extracting long tail data from, 172
 workshops on SEO, 506

X

Xenu Link Sleuth, 513
 XML Sitemaps, 184–187, 267
 file layout, 185
 generators for, 185
 for Google News, 366
 location of file, in robots.txt, 246
 managing and updating, 186
 verification of Sitemaps file, 108
 what to include in file, 186
 where to upload Sitemap file, 186

Y

Yahoo!
 advanced search operators, 59
 combined, 61
 anchor text making pages rank for unrelated terms, 285
 comparison of search results with Google, 481
 Directory Submission Terms, 300
 DMCA notices to, 469
 Local, 340
 News, 363
 reinclusion requests, 466
 robots-nocontent tag, 237
 Search, 499
 linkdomain command, 499
 Search Blog, 474
 Search Content Quality Guidelines, 456
 Search Marketing keyword tool, 148
 search results layout, 24
 SearchMonkey platform, 21
 Site Explorer, 310, 483, 497

- vertical results in SERPs, 27
- vertical search, 66
 - hotjobs, 70
 - News, 68
- Video, 369
- Web Analytics, 117, 377
 - action tracking by referral source, 394
 - conversion tracking, 390, 393
 - most requested pages report, 130
 - unique users report from, 127
- Yellow Pages, 340
 - specialty directories, 342
- YouTube
 - emergence as #2 search engine, 541
 - pros and cons of posting videos on, 370
 - successful use of, 325
 - video search results, 67
 - video upload, 370

ABOUT THE AUTHORS

Eric Enge is the president of Stone Temple Consulting, a leading SEO consulting firm with offices in Boston and northern California. Eric is a regular speaker at search industry conferences, such as Search Engine Strategies (SES), Search Marketing Expo (SMX), and Webmaster World. Eric also makes regular contributions on the topic of SEO to [Search Engine Land](#) and [Search Engine Watch](#), [SEOMoz](#), and his own blogs. Eric is particularly known for his in-depth interviews of major players in the search engine industry, including dozens of people who work in senior capacities at the search engines. Eric is also asked to make guest contributions to *Website* magazine, other SEO-related blogs, and news websites, and is a senior editor for the SEO section of the *Search Engine Marketing Journal*.

Stephan Spencer is founder and president of the highly-respected SEO firm Netconcepts. Its clients include Discovery Channel, HSN, AOL, CNET, and [SuperPages.com](#), to name a few. He is inventor of Netconcepts' automated, pay-for-performance SEO technology platform GravityStream, which powers the natural search channel for online retailers such as Cabela's and Northern Tool. Stephan is a regular contributor to Search Engine Land, Multichannel Merchant, Practical Ecommerce, and MarketingProfs, among others. Stephan is a sought-after conference presenter who's spoken on SEO at hundreds of events across the globe for organizations such as the DMA, AMA, Shop.org, Internet Retailer, SMX, IncisiveMedia (Search Engine Strategies), O'Reilly/TechWeb, PubCon, ECMOD, IQPC, and IIR. Stephan is also an avid blogger. He blogs primarily on his own blog, [Stephan Spencer's Scatterings](#). But his posts can also be found on Searchlight (part of the CNET Blog Network), [Shop.org Blog](#), [NaturalSearchBlog.com](#), [BusinessBlogConsulting.com](#), [MarketingProfs Daily Fix](#), [ChangesForGood.org](#), and [GoogleISuggest.blogspot.com](#).

Rand Fishkin is the CEO and co-founder of [SEOMoz](#), a leader in the field of search engine optimization tools, resources, and community. In 2009, he was named among the 30 Best Young Tech Entrepreneurs under 30 by *BusinessWeek*, and has been written about in the *Seattle Times*, *Newsweek*, and the *New York Times*, among others. Rand has keynoted conferences on search from Sydney to Reykjavik to Montreal to Munich, and has spoken at dozens of shows around the world. He's particularly passionate about the [SEOMoz blog](#), read by tens of

thousands of search professionals each day. In his minuscule spare time, Rand enjoys the company of his amazing wife, Geraldine.

Jessie C. Stricchiola is a search industry veteran and CEO and founder of the highly respected SEO firm Alchemist Media, Inc., consistently recognized by *BtoB Magazine* as one of the top search marketing vendors in the U.S. Jessie lectures extensively on all matters related to SEO and search at leading industry conferences worldwide, including ad:tech, SMX, Incisive Media's Search Engine Strategies, O'Reilly's Web 2.0 Summit, Webmaster World's PubCon, [Shop.org](#), Stanford's Web Publishing Workshop, and others. She is one of the original nine founders of SEMPO (Search Engine Marketing Professional Organization) and served two years on the board of directors for the organization, while chairing the membership committee. Jessie is widely recognized for identifying and publicizing the first case of PPC click fraud in 2001, and her subsequent work to combat the issue has been featured in numerous media outlets, including the *Wall Street Journal, Inc.* magazine, and NPR. Jessie has served as an expert consultant and witness on various Internet- and search-related litigations involving SEO and paid search, as well as patent, trademark, and copyright infringement on the Web.

COLOPHON

The cover image is from *Veer*. The cover fonts are Akzidenz Grotesk and Orator. The text font is Adobe's Meridien; the heading font is ITC Bailey.