

THE EXPERT'S VOICE® IN SHAREPOINT 2010



Pro SharePoint 2010 Search

*Move your company ahead with
SharePoint 2010 search*

Josh Noble, Robert Piddocke,
and Dan Bakmand-Mikalski

Apress®

Pro SharePoint 2010 Search



Josh Noble
Robert Piddocke
Dan Bakmand-Mikalski

Apress®

Pro SharePoint 2010 Search

Copyright © 2011 by Josh Noble, Robert Piddocke, and Dan Bakmand-Mikalski

All rights reserved. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without the prior written permission of the copyright owner and the publisher.

ISBN 978-1-4302-3407-4

ISBN 978-1-4302-3408-1 (eBook)

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

President and Publisher: Paul Manning

Lead Editor: Mark Beckner

Technical Reviewer: Lars Fastrup

Editorial Board: Steve Anglin, Mark Beckner, Ewan Buckingham, Gary Cornell, Jonathan Gennick, Jonathan Hassell, Michelle Lowman, Matthew Moodie, Jeff Olson, Jeffrey Pepper, Frank Pohlmann, Douglas Pundick, Ben Renow-Clarke, Dominic Shakeshaft, Matt Wade, Tom Welsh

Coordinating Editor: Anita Castro

Copy Editor: Mary Ann Fugate

Production Suppoer: Patrick Cunningham

Indexer: John Collin

Artist: April Milne

Cover Designer: Anna Ishchenko

Distributed to the book trade worldwide by Springer Science+Business Media, LLC., 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com.

For information on translations, please e-mail rights@apress.com, or visit www.apress.com.

Apress and friends of ED books may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Special Bulk Sales—eBook Licensing web page at www.apress.com/info/bulksales.

The information in this book is distributed on an “as is” basis, without warranty. Although every precaution has been taken in the preparation of this work, neither the author(s) nor Apress shall have any liability to any person or entity with respect to any loss or damage caused or alleged to be caused directly or indirectly by the information contained in this work.

Contents at a Glance

About the Authors.....	xvi
About the Technical Reviewer	xvii
Acknowledgments	xviii
Introduction	xx
■ Chapter 1: Overview of SharePoint 2010 Search	1
■ Chapter 2: Planning Your Search Deployment.....	23
■ Chapter 3: Setting Up the Crawler	61
■ Chapter 4: Deploying the Search Center.....	109
■ Chapter 5: The Search User Interface.....	121
■ Chapter 6: Configuring Search Settings and the User Interface.....	179
■ Chapter 7: Working with Search Page Layouts	239
■ Chapter 8: Searching Through the API	273
■ Chapter 9: Business Connectivity Services	297
■ Chapter 10: Relevancy and Reporting	359
■ Chapter 11: Search Extensions.....	415
Index	459

Contents

About the Authors.....	xvi
About the Technical Reviewer	xvii
Acknowledgments	xviii
Introduction	xx
■ Chapter 1: Overview of SharePoint 2010 Search.....	1
Microsoft Enterprise Search Products: Choosing the Right Version	2
Microsoft SharePoint Server 2010.....	2
SharePoint Foundation 2010	4
Microsoft Search Server 2010 Express	5
Search Server 2010	5
FAST Search Server 2010 for SharePoint.....	6
Getting to Know Search in SharePoint 2010.....	9
The Search Center	9
Metadata	11
Web Parts	12
SharePoint 2010 Search Architecture	15
Multilingual Support	20
Scaling	21
Extensibility	22
Summary	22

Chapter 2: Planning Your Search Deployment.....	23
SharePoint 2010 Components	24
Web Server Role	24
Query Server Role.....	24
Crawl Server Role	26
Database Server Role	26
Environment Planning and Metrics.....	27
Hardware and Software Requirements	28
Database Considerations: Determining Database Size	30
Initial Deployment Guidelines	33
Typical Server Configurations.....	35
Performance	39
Performance Reports.....	40
Acting on Performance Issues.....	43
Scaling	44
Availability	45
Limitations and Hard-Coded Boundaries.....	47
PowerShell	52
Provisioning Search with PowerShell.....	53
PowerShell for Scaling	58
Summary	59
Further Reading.....	59
SharePoint Components	59
Environment Planning and Metrics.....	59
Performance	59
PowerShell	59

Chapter 3: Setting Up the Crawler	61
The Search Service Application	62
Default Content Access Account	63
Indexing	63
Structured Content	64
Unstructured Content	65
Crawling SharePoint Sites	65
Crawling Users Profiles	67
Synchronizing User Profiles	68
Excluding Directory Tree Nodes	71
Crawling File Shares	72
Crawling Web Sites	74
Crawling Exchange Public Folders	75
Crawling Line-of-Business Data	77
Using Crawl Rules	78
Using Regular Expression in Crawl Rules	80
Using Crawl Rules to Grant Access	83
Troubleshooting Crawl Errors	84
Server Name Mappings	87
Crawler Impact Rules	88
Crawler Scheduling	89
Full vs. Incremental Crawls	90
Crawling Metadata	92
Defining Scopes	94
Adding or Editing Scopes	95
Scope Update Schedule	96
Setting Scopes with PowerShell	96

Crawling Other Document Types with iFilters	98
Adding a File Type to the Content Index.....	98
Federated Sources.....	104
Creating a New Federated Source.....	104
Importing a Federated Location from Microsoft's Federated Search Connector Gallery.....	107
Summary	108
Further Reading.....	108
Indexing	108
Crawl Rules	108
Scopes	108
Federated Sources	108
■ Chapter 4: Deploying the Search Center.....	109
Understanding the Search Center.....	109
Basic vs. Enterprise Search Centers.....	110
The Enterprise Search Center.....	111
Deploying the Enterprise Search Center.....	112
The Basic Search Center	114
Deploying the Basic Search Center	115
Redirecting the Search Box to the Search Center	116
Web Part Deployment	117
Adding Web Parts	118
Summary	119
■ Chapter 5: The Search User Interface.....	121
The Query Box.....	122
Taking Advantage of the Scope Picker	124
The Search Center	126

The Search Results Page	130
Search Suggestions	134
RSS	138
Search Later from Windows Explorer	141
Best Bets	144
Federated Results	145
Search Refiners	149
Search Query Syntax	153
Search Query Operators	154
Property Restrictions	156
The Advanced Search Page	161
Keywords on the Advanced Search Page	162
Picking Property Restrictions	164
Query Examples on the Advanced Search Page	165
The People Search Page	168
People Search Options	169
Standard Web Parts in People Search Results	171
Using People Search Results	171
Taking Action on Results	173
Expertise Search	174
The Preferences Page	176
Summary	177
Chapter 6: Configuring Search Settings and the User Interface	179
Web Parts	180
Web Part Settings	180
Search Box	181
Search Core Results	186

Top Federated Results.....	191
Refinement Panel	193
Federated Search	198
Federated Locations	198
Import Location.....	200
New Location.....	202
Copy Location	207
The Preferences Page: An Administrator's View	208
Understanding Why the Preferences Page Improves Search	208
Preferences Scope	208
Overriding the User's Language Settings	209
Issues to Be Aware Of	209
Stemmers and Word Breakers.....	210
Phonetics and Nicknames	210
Phonetic Search	210
Nickname Search	211
Search Keywords.....	213
Managing Search Keywords.....	214
Maintenance	217
Suggested Uses for Keywords.....	217
Search Alerts Administration.....	218
Search Suggestions.....	219
Viewing Search Suggestions.....	219
Adding Search Suggestions.....	220
Removing Search Suggestions.....	220
Search Scopes.....	221
Managing Scopes	221
Scope Rules.....	224

Display Groups.....	229
Adding Search Scope Selector to the Search Box.....	231
Search Tabs and Pages	232
Summary	237
Chapter 7: Working with Search Page Layouts	239
Understanding Master Pages in SharePoint 2010	240
v4.master	240
default.master	240
minimal.master	240
simple.master	241
Other Master Pages	241
Considerations When Creating a Custom Master Page	242
Adding Navigation to the Search Center	243
Adding Home and Back Buttons to the Search Result Page.....	244
Adding Navigation to Search Center Master Page.....	248
Adding Navigation to the SearchMain.aspx Page.....	250
Applying a Branded Master Page to a Search Center.....	252
Search Box Is Hiding When Branding a Search Center	252
Making Navigation Visible and Showing the Search Box Correctly	253
Adding CSS Styling.....	254
Modify the Search Results Presentation.....	256
Adding Additional Info to Search Results with XSLT	256
Enabling Ratings.....	256
Making Ratings Searchable.....	257
Adding Ratings to Search Results	259
Viewing and Searching for Ratings	260
Removing Metadata from Search Results	262

Changing Click Action of Search Results.....	262
Creating a Managed ID Property.....	263
Advanced Topics on Refinement Panel	265
Adding Refiner Counts.....	265
Changing the Accuracy Index	267
Number of Categories to Display	269
Metadata Threshold.....	270
Number of Characters to Display.....	270
Summary	271
Chapter 8: Searching Through the API	273
Understanding the Query Expression Parsing Order.....	273
Operator Order.....	274
Using a Tree Structure to Understand the Query.....	275
Manipulate the Query with Parentheses.....	276
The Search API	276
Creating a KeywordQuery-Based Search	277
Creating a FullTextSqlQuery-Based Search.....	278
Searching Through the Search Web Service	278
Creating SQL for the FullTextSqlQuery	282
Creating the Query.....	282
Important Configuration Options of the Query Object.....	282
SQL Query Predicates	282
Details on Keyword Inclusion	283
FREETEXT and CONTAINS Predicates Format.....	284
Creating a Custom Search Application Page	286
Setting Up the Solution	287
Modifying the Application Page Markup.....	290
Adding Code-Behind.....	292

Summary	295
■ Chapter 9: Business Connectivity Services	297
BCS Architecture.....	298
Presentation	298
Core Components	301
Content Sources	302
Business Connectivity Services Deployment Types.....	302
Functionality for Normal Users “Out-of-the-Box”	302
Functionality for Advanced Users and Administrators.....	303
Code-Based Solutions	304
Set the Title Field in the External Content Type.....	312
Creating a Profile Page to Display BCS Results.....	313
Configure Indexing.....	317
Performing a Search.....	318
Creating a .NET Connector in Visual Studio 2010	320
Creating a Sample Flat File Data Source	320
Creating a .NET Assembly Connector Project in Visual Studio 2010	320
Creating an Entity (External Content Type)	323
Creating an Entity Service Class.....	324
BDC Modeling Tools.....	325
Defining the BDC Model.....	328
Deployment	335
The Secure Store Service	335
Configuring the Secure Store Service	335
Creating a Secure Store Service Application for Impersonating	339
Setting the Application Impersonation Credentials	344
Setting Permissions.....	346

Creating Security Trimmed CRUD Operations on a SQL Database Using Visual Studio 2010	348
Connecting Model to Data Source	349
Mapping BDC Model to Data Source.....	350
Adding Code-Behind to Access External Data Source	354
Adding Security Trimming to .NET Connectors.....	355
Summary	358
Chapter 10: Relevancy and Reporting	359
Relevancy Algorithms	360
Keyword Matches.....	361
Proximity	363
Static Document Relationships.....	363
User-Driven Weighting	364
The Corpus.....	365
Search Reporting	366
Farm Level Reports.....	367
Site Collection Reports	369
Number of Queries.....	372
Top Queries	372
Failed Queries.....	373
Best Bet Usage	375
Best Bet Suggestions	375
Best Bet Suggestions Action History	375
Enabling Reporting	375
Displaying Report Data in Site	376
Tuning Search.....	377
Authoritative Pages	378
Result Removal.....	380

Stop Words	381
The Thesaurus	382
Custom Dictionaries	385
The noindex Class.....	387
Popularity (Click-Through)	388
Social Tagging	388
The Ratings Column.....	388
Search Keywords and Best Bets	396
Managed Metadata Service	399
Tagging	403
Custom Ranking.....	406
Built-In Models	406
Custom Models.....	408
Setting the Results Webpart to Use Custom Ranking Models	412
Summary	414
Further Reading	414
Relevancy	414
Reporting	414
Tuning Search	414
Custom Ranking	414
■ Chapter 11: Search Extensions.....	415
Extending Search Beyond In-House Development	416
CodePlex and Open Source.....	416
Getting the Most Out of Your Open Source Products.....	418
Additional Search Related Projects	418

Commercial Solutions.....	421
User Interface Enhancements	421
Comprehensive Commercial Search Extensions	424
Vendor Notes	430
Visualization	431
Scanning Options.....	432
Lightweight Document Viewers.....	432
Other Options.....	433
Choosing the Best Option	433
iFilters	435
Taxonomy, Ontology, and Metadata Management	438
Automated Classification.....	439
conceptClassifier	439
Choosing the Best Option	443
Replacing the SharePoint Search Engine	445
Replacement Considerations.....	446
Google Search Appliance.....	447
FAST Search Server 2010 for SharePoint.....	450
Choosing the Best Option	455
Summary	458
 Index	 459

About the Authors



■ **Josh Noble** has made a career on the front lines, diagnosing business pains and alleviating them with technology. As the sales manager for SurfRay Inc., Josh has worked with global clients, such as the American Bar Association, Best Buy, Intel, the US Coast Guard, and Conoco Phillips, to build solutions to business problems through expertly designed search on Microsoft CMS and SharePoint. His unique background allows him to assist teams to bridge the gap between the business unit's project goals and the technical unit's project design. Josh is from Fort Worth, TX, where he received his Bachelor of Science degree from Texas Christian University. He is currently studying for his MBA at his alma mater. Josh can be contacted on e-mail at josh@prosharepointsearch.com.



■ **Robert Piddocke** has worked with search technology for over a decade and has implemented hundreds of site and enterprise search solutions, including solutions for the Holy See, the London Stock Exchange, Lloyds of London, Microsoft Xbox, MetLife, and many more. Robert has worked for several search vendors and implemented a range of enterprise and site class search products. Currently, Robert is technical lead in North America for SurfRay Inc., the leading search add-on vendor for SharePoint. Robert has lived and worked in Japan and Denmark and speaks English, Danish, and Japanese. He is originally from Vancouver, BC, where he now lives with his wife and two children. He has his degree from the University of Victoria, in Victoria, BC. Robert can be reached at Robert@prosharepointsearch.com.



■ **Dan Bakmand-Mikalski** works with research and development at the global Danish company SurfRay Inc., a market leader in search technology on the SharePoint platform. First-hand experience with customers and SharePoint professionals has given him a unique insight into the core challenges you are likely to face when doing advanced customizations and extensions to native SharePoint search. Specializing in understanding the SharePoint search engine characteristics and the search interfaces, his experience offers relevant insights for newcomers as well as experienced SharePoint administrators and developers on how to leverage unused potential in SharePoint. Dan has an academic background as a civil engineer in the fields of computer science and mathematical modeling from the Technical University of Denmark. Dan was born in Copenhagen, Denmark in 1978, where he also currently lives and works.

Although Dan has written several articles on SharePoint and search technology, this book is Dan's debut as a published author and professional writer. Dan can be contacted on e-mail at bakmand@gmail.com.

About the Technical Reviewer



■ **Lars Fastrup** is an independent contractor offering software development and advisory services for SharePoint products and technologies, the .NET Framework, and Microsoft SQL Server. Until the end of 2007, he worked for the search technology company Mondosoft, where he headed the conception and development of the popular Ontolica search add-ons for SharePoint Portal Server 2003 and SharePoint Server 2007. Lars also co-authored the book *Inside the Index and Search Engines: Microsoft Office SharePoint Server 2007*. Besides his consulting business, Lars also develops and sells add-on products for SharePoint from his web site, www.sharepointproducts.com.

Acknowledgments

When Apress editor Mark Beckner reached out nine months ago, asking me to bring a team together to write this book, I thought he must have lost his mind. Matt, I would like to thank you for placing your faith in us. Without your drive to convince me that we were capable of taking on such an ambitious project, none of this would have been possible.

I would like to thank my co-authors, Dan and Robert. The time and dedication you invested into not only writing this book but also putting up with me should be considered saintly. Dan, I don't know how you survived so many late nights still working in Denmark past the hours I fell asleep in the United States. Robert, I would like to thank you for never ignoring me when I wanted to chat about the latest new idea or vent about problems with our test server. It has been an amazing journey uncovering all the hidden valleys of search in SharePoint with you, and I could not have asked for a better team.

They say that behind every great man there is a great woman. While I am not yet a great man, I am fortunate to be halfway there with a great woman. To my amazing fiancée, I want to thank you for placing me on loan to this project for the past several months. For all the late nights and exhausting weekends researching and writing, your support grounded my sanity. Thank you for always pretending to care when I rambled about the latest feature we found or challenge we encountered.

To my family, friends, and colleagues, thank you for supporting me over these many months. I would like to thank my father for instilling my work ethic, exposing me to technology from a young age, and always pushing me to seek the next challenge. To my mother, I would like to thank you for providing me with a good upbringing while striving to make me compassionate, dedicated, ethical, and humble. For my many supportive friends and MBA team, I thank you for your constant encouragement. To my partners, clients, and peers, thank you for always asking the hard questions and challenging me to learn.

To our technical editor, Lars, I want to thank you not only for correcting us when we were wrong and pointing out everything we missed, but also for writing the amazing resource that this book built upon. Without your book, *Inside the Index and Search Engines*, we would still be planning our table of contents. To our colleagues at SurfRay and CEO Soren Pallesen, I would like to thank you for allowing us to share our knowledge of SharePoint search with the world, even when doing so was uncomfortable. Finally I would like to thank our coordinating editor, Anita Castro, development editor, Jim Markham, and marketing manager, Leo Cuellar, for your dedication to this book. If it weren't for the support of everyone around me, this book would have never been possible.

Josh Noble

Writing this book was a lot of fun. It was also a lot of work, work that was done during time I should have been paying attention to my wife and kids. So I want to thank Maya, Pavel, and Joanna, first and foremost, for supporting me for the six months of evenings I spent writing, testing, and editing, and not with them. Additionally, special thanks go to my good friend and our technical editor, Lars Fastrup, for his guidance, support, and great and helpful edits. Thanks also to the team at Apress, including Anita, Jim, Leo, and Mark for picking us up, dusting us off, and keeping us on track. Finally, thanks to my great co-authors, each with their own style: Josh, who kept us on track and hassled us when deadlines approached, and Dan, for his technical insight and fun attitude. Thanks guys! A better team I could not have hoped for. I hope this book is as much fun to read as it was to write and useful for those interested in getting the most out of SharePoint Search.

Robert Piddocke

Special thanks go to Robert and especially Josh for their work in managing the editorial and publishing process and their gentle reminders and encouragement during the busy days and nights of writing. Special appreciation goes toward my family and friends, whom I have seen much less of during the writing process than I would like. Thank you for showing great interest and understanding during a busy period. It has been a great motivation.

Dan Bakmand-Mikalski

Introduction

Why Is This Book Useful?

This book has been written to address what no other single resource has been dedicated to tackle, search in SharePoint 2010 (SPS 2010). While there are other books that spend a brief chapter to touch on search in SharePoint 2010, scattered information in Microsoft documentation and on blogs, and SharePoint search books that actually focus more on FAST Search Server 2010 for SharePoint than SharePoint's own search capabilities, at the time of this book's publication, there are no other books devoted explicitly to the search offering included in SharePoint 2010. General SharePoint resources may spend 50 pages summarizing the Microsoft documentation on search, but they cannot do more than scratch the surface in such an abbreviated space. Other search-focused books explain the theoretical concepts of enterprise search, or jump heavily into Microsoft's new product, FAST Search Server 2010 for SharePoint. This book, by contrast, is beneficial to all deployments of SharePoint 2010. The information presented throughout is applicable to standard and enterprise editions of the platform. Due to the great amount of overlap, it is also widely useful for deployments of Search Server 2010 and Search Server 2010 Express.

While there are many technical resources about SharePoint 2010 available that were produced with Microsoft oversight, this is not one of them. As a result, this book is able to dive into the hard-to-find details about search in SharePoint 2010 that are not widely exposed. We hope this book will help teach you how to do what consultants charge a fortune to do, and help you understand the best way to do it. We share our years of experience maximizing SharePoint and other enterprise search engines. We not only take a look inside the machine and show you the gears, but also explain how they work, teach you how to fix the problem cogs, and help you add efficiency upgrades.

This book is an end-to-end guide covering the breadth of topics from planning to custom development on SPS 2010. It is useful for readers of all skillsets that want to learn more about the search engine included in SharePoint 2010. After reading this book, you will be able to design, deploy, and customize a SharePoint 2010 Search deployment and maximize the platform's potential for your organization.

Who Is This Book Written for?

Quite a bit of energy was put into insuring this book is useful for everyone with an interest in SharePoint 2010 Search. It was purposefully written by a SharePoint developer, a SharePoint administrator, and a business consultant so that each could contribute in his respective areas of expertise. The chapters have been designed to evenly cater to three primary readers: users, administrators, and developers.

We recognize that most readers will not utilize this book cover to cover. To make it more useful for the varying areas of interest for reader groups, instead of meshing topics for various groups into each chapter, we have designed the chapters to primarily touch on topics for one reader group. For example, Chapter 5 was written to teach users about using the search user interface, Chapter 10 sticks to the administrator topic of utilizing farm analytics to improve search relevancy, and Chapter 9 teaches

developers how to build custom connectors for the BCS. No matter your level of expertise, there are topics in this book for anyone with an interest in getting the most out of search in SharePoint 2010.

The following are some of the key topics throughout the book that will be useful for readers with various needs.

Topics for Users

- *Components of the search interface:* Chapter 5 provides a thorough walkthrough of the various components of the search interface, including the locations of features and how they work.
- *Setting alerts:* Chapter 5 explains alerts and provides a guide on how to use and set them.
- *Query syntax:* Chapter 5 provides a full guide to the search syntax, which can be used in query boxes throughout SharePoint to expand or refine searches.
- *Using the Advanced Search page:* Chapter 5 outlines the Advanced Search page and how it can be used to expand and scope queries.
- *Using people search:* Chapter 5 teaches the components of the people search center and how to use the people search center.
- *Using the Preferences page:* Chapter 5 explains when the Preferences page should be used and how to use it.

Topics for Administrators

- *Managing the index engine:* Chapter 3 goes into detail on setting up the crawler for various content sources, troubleshooting crawl errors, and using iFilters.
- *Deploying search centers:* Chapter 4 explains the techniques and considerations for deploying search centers.
- *Configuring the search user interface:* Chapter 6 builds on Chapter 5 by providing a detailed walkthrough on configuring search Web Parts, search centers, and search-related features.
- *Setting up analytics and making use of analytical data:* Chapter 10 focuses on the setup of SharePoint reporting and using the data to improve business processes and relevancy.
- *Tuning search result relevancy:* Chapter 10 provides detailed instruction regarding how to improve search result relevancy by using features such as authoritative pages, synonyms, stop words, the thesaurus, custom dictionaries, ratings, keywords, and best bets.
- *Managing metadata:* Chapter 10 dives into the uses of metadata in SPS 2010 Search, how to set up metadata, and how to use it to improve relevancy of search results.

- *Creating custom ranking models*: Chapter 10 ends by covering the advanced topic of utilizing PowerShell to create and deploy custom relevancy ranking models.
- *Enhancing search with third-party tools*: Chapter 11 discusses commercial third-party tools that enhance search beyond functionality available with light custom development.

Topics for Developers

- *Adding custom categories to the refinement panel Web Part*: Chapter 6 discusses the most essential search Web Part customizations, including how to add new refinement categories to the refinement panel Web Part.
- *Designing custom search layouts*: Chapter 7 covers subjects necessary to design a search interface with a custom look and feel. Topics necessary for this include manipulation of master pages, CSS, and XSLTs.
- *Modifying the search result presentation*: Chapter 7 provides instruction for changing result click actions and editing the information returned for each search result with XSL modifications.
- *Improving navigation in search centers*: Chapter 7 gives detailed instruction for adding site navigation to the search interface, which is disabled by default.
- *Advanced customization of the refinement panel Web Part*: Chapter 7 provides instruction for advanced customization of the refinement panel Web Part.
- *Creating custom search-enabled applications*: Chapter 8 covers topics such as the search API and building custom Web Parts with Visual Studio 2010.
- *Creating Business Connectivity Services components*: Chapter 9 exclusively covers end-to-end topics on connecting to external content sources through the Business Connectivity Services (BCS).

What Topics Are Discussed?

This book covers the end-to-end subject of search in SharePoint 2010. We start with a brief background on the available Microsoft search products and follow with key terms and a basic overview of SPS 2010 Search. The book then guides readers through the full range of topics surrounding SharePoint search. We start with architecture planning and move through back-end setup and deployment of the search center. We then jump into an overview of the key user-side features of search, followed by how to configure them. More advanced topics are then introduced, such as custom development on the user interface, leveraging the BCS to connect to additional content sources, and how to use search analytics to improve relevancy. The book is capped off with a chapter on how improve search beyond the limitations of the base platform.

While this provides a general overview of the path of the book, each chapter contains several key topics that we have found to be important to fully understand SharePoint 2010 Search from the index to the user experience. These are the key concepts learned in each chapter.

Chapter 1: Overview of SharePoint 2010 Search

This chapter introduces readers to search in SharePoint 2010. It provides an overview of the various Microsoft search products currently offered and their relation to each other as well as this book. A brief history of SharePoint is given to explain developments over the last decade. The chapter lays the groundwork of key terms that are vital to understanding search in both SharePoint and other search engines. It explains the high-level architecture and key components of search in SPS 2010. It also provides a guide for topics throughout the book that will be useful for various readers.

Chapter 2: Planning Your Search Deployment

This chapter provides further details of the core components of SharePoint 2010 Search, and issues that should be taken into account when planning a deployment. Each component of search and its unique role are explained at further length. The function of search components as independent units and a collective suite is addressed. Hardware and software requirements are outlined, and key suggestions from the authors' experience are given. Scaling best practices are provided to help estimate storage requirements, identify factors that will affect query and crawl times, and improve overall search performance. Redundancy best practices are also discussed to assist in planning for availability and avoiding downtime.

Chapter 3: Setting Up the Crawler

This chapter dives into setup of the index engine and content sources. It provides step-by-step instructions on adding or removing content sources to be crawled as well as settings specific for those sources. It covers how to import user profiles from Active Directory and LDAP servers and index those profiles into the search database. Crawling and crawl rules are addressed, and guidance on common problems, including troubleshooting suggestions, is given. The chapter also explains how crawl rules can be applied to modify the connection credentials with content sources. Finally, the chapter explains the setup of iFilters to index file types not supported out of the box by SharePoint 2010.

Chapter 4: Deploying the Search Center

This brief chapter provides step-by-step instructions on deploying SharePoint search centers. It explains search site templates and the difference between the two options available in basic SPS2010. A guide on redirecting the search box to a search center is given, as well as notes on how to integrate search Web Parts into sites other than the search center templates.

Chapter 5: The Search User Interface

This chapter is an end-to-end walkthrough of the search user interface in SPS2010. A wide range of topics is discussed to provide a comprehensive user guide to search. It explains how to use the query box and search center to find items in SharePoint. It explains the different features of SharePoint search that are accessible to users by default, such as the refinement panel, alerts, and scopes. A full guide on search syntax is given for advanced users, and a guide of the people search center is provided for deployments utilizing the functionality.

Chapter 6: Configuring Search Settings and the User Interface

This chapter expands on Chapter 5 by diving into configuration of the search user interface. It provides advice on how to accomplish typical tasks for configuring the search user experience in SPS 2010. The first part of the chapter explains the common search Web Parts and their most noteworthy settings. The following parts of the chapter focus on understanding concepts such as stemmers, word breakers, and phonetic search. The chapter provides details on configuring general search-related settings such as scopes, keywords, search suggestions, refiners, and federated locations. Information on administrative topics related to user settings, such as search alerts and user preferences, is also described in detail.

Chapter 7: Working with Search Page Layouts

This chapter is the first of two that focus on advanced developer topics related to search. It explains best practices for design and application of custom branded layouts to the search experience. Topics such as manipulation of the CSS, XSLTs, and master pages are all specifically addressed. A detailed discussion of improving navigation within the search center is also provided. The chapter continues with guidance on manipulating the presentation of properties and click action of search results. It ends with instruction for advanced customization of the refinement panel Web Part.

Chapter 8: Searching through the API

This is the second of two chapters that focus on advanced developer topics related to search. It delivers the fundamentals of the search application programming interfaces (APIs) in SharePoint 2010. A thorough re-introduction to the query expression is presented from a development perspective, and guidance is provided on how to organize the query expression to get the desired results. The chapter also contains an example of how to create a custom search-enabled application page using Visual Studio 2010.

Chapter 9: Business Connectivity Services

This chapter is an end-to-end guide for developers on the SharePoint 2010 Business Connectivity Services (BCS) with a special focus on the search-related topics. It explains the architecture of this service and how it integrates both within and outside SharePoint 2010. A guide is given on how to create BCS solutions and protocol handlers, including a full step-by-step example. Specific examples are also provided of how to use SharePoint Designer 2010 to create declarative solutions and Visual Studio 2010 to create custom content types using C#.

Chapter 10: Relevancy and Reporting

This chapter is a guide for the user of SharePoint analytics and applications to improve search relevancy. It teaches readers how to view and understand SharePoint search reporting and apply what it exposes to enhance the search experience. A guide to the basics of search ranking and relevancy is provided. The key settings that can be applied to manipulate items to rise or fall in search results are explained. Reporting and its ability to expose the successes and failures of the search engine are explained, along with techniques that can be applied to modify the way the search engine behaves. A guide to utilizing the SharePoint thesaurus to create synonyms for search terms is also provided. The chapter ends with advanced instructions for utilizing PowerShell to create and deploy custom ranking models.

Chapter 11: Search Extensions

This chapter explains the limitations of SharePoint 2010 and various options for adding functionality to the platform beyond custom development. It is the only chapter that explores topics beyond the capabilities of the base platform. It explores the business needs that may require add-on software, and reviews vendors with commercial software solutions. It takes a look into free add-on solutions through open source project communities, and provides general outlines of when replacements to the SharePoint 2010 Search engine, such as FAST Search Server for SharePoint 2010 (FAST) or Google Search Appliance, should be considered.

This Is Not MOSS 2007

While skills picked up during time spent with MOSS 2007 are beneficial in SPS 2010, relying on that expertise alone will cause you to miss a lot. There have been significant changes between MOSS 2007 and SharePoint 2010. Search not only received improvement, but also underwent complete paradigm shifts. The old Shared Services Provider architecture has been replaced with the SharePoint 2010 service application architecture, creating unique design considerations. The MOSS 2007 Business Data Catalog (BDC) has been replaced with the Business Connectivity Services (BCS), unlocking new ways to read and write between SharePoint and external content sources. Index speed, capacity, and redundancy options have all been improved to cater to expanding enterprise search demands. Even the query language has been completely revamped to allow for Boolean operators and partial word search.

Throughout this book, we have taken special care to note improvements and deviations from MOSS 2007 to assist with learning the new platform. Captions pointing out changes will help you to efficiently pick up the nuances of SharePoint 2010. Direct feature comparisons are also provided to assist with recognizing new potential opportunities for improving search.

The Importance of Quality Findability

If you are reading this book, then most likely your organization has decided to take the leap into SharePoint 2010. Unfortunately, more often than not the platform is selected before anyone determines how it will be used. This leaves a large gap between what the platform is capable of achieving and what is actually delivered to users. The goal of this book is to bridge the gap between what SharePoint can do to connect users with information, and what it does do for your users to connect them with their information.

By default, most of the world's computer owners have a browser home page set to a search engine. Search is the first tool we rely on to find the needle we need in a continuously expanding haystack of information. People expect search to quickly return what they are looking for with high relevancy and minimal effort. Improvements catering to effective Internet search have raised user expectations, which should be seen as a call to action for improved web site and portal design, not an opportunity to manage expectations. If this call to action is not met, however, business will be lost to completion for web sites, and intranet users will find shortcuts to the desired content management practices.

Consider your own experiences on your favorite global search engine. If the web site you are looking for does not appear within the first (or maximum, second) page of search results, then you most likely change your query, utilize a different search engine, or simply give up. Users on SharePoint portals exhibit the same behavior. After a few attempts to find an item, users will abandon search in favor of manual navigation to document libraries or the shared drives that SharePoint was designed to replace. Users eventually begin to assume that once items find their way into the chasm of the intranet, the only chance of retrieving them again is to know exactly where they were placed. It is for these reasons that

implementing an effective search experience in SharePoint 2010 is one of the most important design considerations in SharePoint. If users cannot easily find information within your SharePoint deployment, then they cannot fully leverage the other benefits of the platform.

The Value of Efficient Search

It is obvious that in today's economy it is more important than ever to make every dollar count. Organizations cannot sit back and ignore one of the largest wastes of man-hours in many companies. According to a 2007 IDC study, an average employee spends 9.5 work-hours a week just searching for pre-existing information. What's worse is that six hours a week are spent recreating documents that exist but cannot be found. With this information, combined with the statistic that users are typically successful with their searches only about 40% of the time, the cost of a poor search solution can quickly compound to quite a large burden on a company of any size.

Let's say that an employee is paid \$75,000 a year for a 40-hour work week and 50 weeks a year (2,000 hours). Based on this, the employee earns \$37.50/hour before benefits. Applying the statistics just cited, you can see that the cost per week to find information is \$337.50/week (\$16,875 annual), and the cost to recreate information is \$225.00/week (\$11,250 annual). This being said, the cost per employee at this rate would be \$28,125/year for a poor findability and search solution. In a different deployment scenario, assume 500 employees earning \$20 per hour, with just one hour loss per user/month. In just three months, the waste due to poor search is \$30,000 in wasted wages. That is an extra employee in many companies.

From these statistics, it is clear that well-designed search is a key driver of efficiency within companies. This book helps you to achieve this efficiency with search. It provides a full range of topics to help you design a SharePoint search portal that quickly connects users with their information. We pull from our experience working with SharePoint search every day to provide expert advice on the topics that matter when building a SharePoint search center that really works. Although designing and implementing a quality search experience does take time, this book places the ability within the grasp of every SharePoint 2010 deployment.

Note from the Authors

Our goal is not only to teach you the facts about search in SharePoint 2010, but also to give you the basic tools to continue your learning. Creative applications for SharePoint search are always evolving. Use the knowledge gained in this resource to explore the continuing evolution of knowledge throughout your company, peers, and the Web. As you build your SharePoint search environments, make sure to always keep the users' experiences in mind. Solicit feedback, and continue to ask yourself if the search tool you are creating will help users change *search* into *find*.

This book is the product of countless hours of planning, research, and testing. It is the combined efforts of many people, including Apress editors, Microsoft, SharePoint consultants, bloggers, clients, and our colleagues at SurfRay. With these people's support, we have designed this book's content and structure to teach you all the essentials of search in SharePoint 2010. As you continue on to Chapter 1, we hope that you enjoy reading this book to the same extent we have enjoyed writing it for you.



Overview of SharePoint 2010 Search

After completing this chapter, you will be able to

- Distinguish between the various Microsoft Search products
- Understand the search architecture in SharePoint 2010
- Translate integral terms used throughout the rest of the book
- Know how to effectively use this book

Before taking the journey into this book, it is vital to gain a firm understanding of the ground-level concepts that will be built upon throughout. This chapter is designed to bring together several of the core concepts necessary to understand the inner workings of SharePoint 2010. Many of these are universal to all search engines, but some may be foreign to those readers new to SharePoint.

It is important to keep in mind that a few of the terms used throughout this resource may be different than those used on public blogs and forums. The terminology presented in this chapter will assist the reader in understanding the rest of this book. However, it is more important to understand the core concepts in this chapter as they will prove more helpful in your outside research. As discussed in the introduction, this book will not address every possible topic on search in SharePoint 2010. The most important subjects are presented based on the experiences of the authors. The dynamics of SharePoint, however, create a potentially unending network of beneficial topics, customizations, and developments. While this book does not cover everything, it will provide all of the basic knowledge needed to effectively utilize additional outside knowledge.

Microsoft has a wide range of enterprise search product offerings. With new products being released, and existing products changing every few years, it can become quite cumbersome to keep track of new developments. To lay the foundations of the book, the chapter starts with a brief review of this product catalog. Each solution is explained from a high level with specific notes on the key benefits of the product, technological restrictions, and how it fits into this book. While it is assumed that every reader is using SharePoint 2010, a large amount of the topics discussed will be relevant to other products in the Microsoft catalog.

The second half of the chapter first focuses on a few of the most important soft components of search. These include components such as the search center, the document properties that affect search, and the interactive components for users. The second half of the chapter then outlines the basic architecture of SharePoint 2010 Search. While this topic is discussed at length in the following chapter,

the depth of detail provided here is sufficient for readers not involved with the infrastructure setup. Finally, the chapter is capped with a guide to a few of the most important topics in this book for various reader groups.

Microsoft Enterprise Search Products: Choosing the Right Version

As mentioned in the introduction, Microsoft has been in the search space for over a decade. In that time, they have developed a number of search products and technologies. These range from global search on Bing.com, desktop search on Windows 7, search within Office 14, and a wide range of “enterprise” search solutions. Each of these products is designed to handle specific types of queries, search against various content sources, and return results using various ranking algorithms. No two search technologies are the same, and a user being fluent in one does not translate to effective use or deployment of another. For the purpose of this book, we will be focusing on Microsoft SharePoint 2010, and as the weight of this book indicates, this subject is more than enough information for one resource.

Due to the overlap between many of Microsoft’s enterprise search technologies, we will make side notes throughout this book indicating where the information is applicable to solutions other than SharePoint 2010. Throughout the book there will also be notes on technology limitations, where the use of an additional Microsoft technology or third-party program may be necessary to meet project goals. These side notes should not be considered the definitive authority on functionality outside the scope of this book, but they are useful in recognizing key similarities and differences between products.

Microsoft SharePoint Server 2010

SharePoint Server 2010 is Microsoft’s premier enterprise content management and collaboration platform. It is a bundled collection of workflows, Web Parts, templates, services, and solutions built on top of Microsoft’s basic platform, SharePoint Foundation, which is discussed further in the following section. SharePoint 2010 can be used to host a wide variety of business solutions such as web sites, portals, extranets, intranets, web content management systems, search engines, social networks, blogs, and business intelligence databases.

SharePoint 2010 deployments can be found in organizations with a massive difference in scale and requirements. User counts in implementations as small as single digits are seen in small intranets and expand into the millions with large extranets and public-facing sites. The beauty of the solution comes in its ability to be deployed relatively quickly and easily, and its ability to be customized to cater to a wide range of needs with various workflows, Web Parts, templates, and services. The out-of-the-box functionality can cater to generic needs of organizations, but the power of the tool comes in the building blocks that are able to be inserted, combined, and customized to meet a variety of usage scenarios. While the most obvious use of SharePoint 2010 is intranet portals, the platform is now seeing a greater push to the public domain with wider-range Web 2.0–focused tools.

SharePoint 2010 is available both on-premise, off-site, and in the cloud through Microsoft as well as several third-party hosting firms. On-premise refers to deployments of software that run locally on in-house hardware, as opposed to those that are hosted in a remote facility, such as a server farm or on the internet. Historically, most software has been managed through a centralized on-premise approach, but in recent years, advances in cloud computing, the rise of netbooks, and the availability of inexpensive broadband have grown the popularity of decentralized off-premise deployments. While both approaches can produce the same experience for users, each presents its own set of IT challenges. On-premise deployments require the procurement, maintenance, upgrade costs, and potential downtime

associated with server hardware. Off-premise deployments at hosting centers allow companies to avoid these challenges for a fee, but present their own challenges in the way of bandwidth, security, and more limited functionality depending on the hosting center. Off-premise options for SharePoint 2010 are available through various hosting centers. Many of these hosts simply maintain reliable off-site deployment of the same software available internally and provide remote access to full configurability options. Other hosted versions, such as SharePoint Online offered by Microsoft, may provide only a subset of the features available through on-premise deployments. Due to the variable features available in the off-premise offerings, this book will target the on-premise version of SPS 2010.

Unlike SharePoint Foundation 2010, which will be discussed in the next section, SharePoint Server 2010 requires additional software licensing. Licensing costs may deviate depending on a particular client's licensing agreement and procurement channel. Microsoft may also deem it necessary to change licensing structures or costs from time to time. As a result, this book will not discuss licensing costs, although this should be taken into consideration during the planning stages discussed in Chapter 2.

Before learning about the current version of SharePoint, it may be helpful to know the background of products it has been derived from. SharePoint 2010 stems from a decade and a half of development history. During this time, Microsoft has taken note of the platform's pitfalls and successes to continuously produce improved platforms every few years. Fueled by the need to be able to centrally share content and manage web sites and applications, the earliest version of SharePoint, called Site Server, was originally designed for internal replacement of shared folders. Site Server was made available for purchase with a limited splash in 1996 with capabilities around search, order processing, and personalization.

Microsoft eventually productized SharePoint in 2001 with the release of two solutions, SharePoint Team Services (STS) and SharePoint Portal Services (SPS 2001). SharePoint Team Services allowed teams to build sites and organize documents. SharePoint Portal Services was focused primarily on the administrator and allowed for structured aggregation of corporate information. SPS also allowed for search and navigation through structured data. Unfortunately, the gaps between these two solutions created a disconnect between the end users using SharePoint Team Services to create sites and administrators using SharePoint Portal Services to manage back-end content.

In 2003, Microsoft released the first comprehensive suite that combined the capabilities of SharePoint Team Services and SharePoint Portal Services. Much like today, the 2003 version of SharePoint came in two different flavors, Windows SharePoint Services 2.0 (WSS 2.0), which was licensed with Windows Server, and SharePoint Portal Server 2003 (SPS 2003). Due to the inclusion of WSS 2.0 in Windows Server, and the large improvements over the 2001 solutions, adoption of SharePoint as a platform began to skyrocket. SharePoint 2003 included dashboards for each user interface, removed much of the tedious coding required in previous versions, and streamlined the process for uploading, retrieving, and editing documents.

In 2006, Microsoft released Microsoft Office Server 2007 (MOSS 2007) and Windows SharePoint Services 3.0 (WSS 3.0), following the same functionality and licensing concepts of their 2003 counterparts. By leveraging improvements in the underlying framework, SharePoint 2007 ushered in the maturity of the platform by introducing rich functionality such as master pages, workflows, and collaborative applications. MOSS 2007's wide range of improvements from administrative tools to user interfaces positioned SharePoint as the fastest growing business segment in Microsoft.

In May 2010, Microsoft released SharePoint Server 2010 (SPS 2010) and SharePoint Foundation 2010, the successors to MOSS 2007 and WSS 3.0. SharePoint 2010 builds on MOSS 2007 by improving functions such as workflows, taxonomy, social networking, records management, and business intelligence. It is also noteworthy to point out Microsoft's noticeable improvements to features catering to public-facing sites and cloud computing.

In regards to search, improvements in SharePoint 2010 can be found across the board in areas such as improved metadata management, the ribbon, inclusion of the Business Connectivity Services (BCS) in non-enterprise versions, a significantly more scalable index, expanded search syntax, and search refiners (facets). With the exception of metadata management, these are the types of subjects that will be