

CONTENTS

INTRODUCTION

xxvii

PART I: INTRODUCTION

CHAPTER 1: WHAT IS ENTERPRISE SEARCH?	3
Why Enterprise Search?	4
How Enterprise Search Differs from Web Search	4
The Enterprise Is Not Just a Small Internet	4
Technical Differences in Search Requirements and Technologies	5
Every Intranet Search Project Is Unique	6
Enterprise Search Technology Overview	6
Search Components Outline	6
Vendor Vocabulary	7
Scalability	8
Federated Search	8
Microsoft's 2010 Search Technology Road Map	8
Microsoft's Explanation	9
Categorizing Your Organization's Use of Search — Examples	9
The Gray Area: Where Search Might Be Misranked	10
A Local Sales Office	10
A Contract Manufacturer	11
A Midsized to Large Energy Producer	11
The ROI of Search	11
The Business Intelligence Benefits of Search	12
The Search Center of Excellence	14
Before SCOE	14
The Birth of SCOE	14
Centralized SCOE	15
Distributed SCOE	15
Examples of SCOE Tasks	15
SCOE Staffing and Skills	16
Questions That Everybody Asks	16
Why Not Just Use an Open-Source Engine Like Lucene?	18
I Have Search Inside SharePoint; Do I Need Anything Else?	19

Why Is Search so Hard?	19
We've Upgraded Search, and Users Are Unhappy.	20
Do I Need a Taxonomy?	21
Do I Need Facets or Entity Extraction?	22
What Is Clustering?	23
Can't I Just Have Users Tag Everything?	23
Can We Do It in Phases?	24
How Much Is This Going to Cost?	24
What Type of Staffing Do I Need for Search?	25
Isn't Enterprise Search Dead?	26
How Important Is the Gartner Magic Quadrant?	26
Summary	27
CHAPTER 2: DEVELOPING A STRATEGY — THE BUSINESS PROCESS OF SEARCH	29
Developing the Search Project Plan	30
eCommerce Search	30
Intranet and Extranet Search	31
File Share Search	32
Public Site Search	33
Organizing for Success: the Search Center of Excellence	34
Centers of Excellence	34
Building a Virtual SCOE	35
The Process of Acquiring a Search Platform	35
Building a Business Case	35
Hard ROI	36
The Soft ROI	37
Defining Your Requirements	38
An Overall Approach	38
High-Level Questions	39
Stakeholders	44
The Role of Analysts	49
Priorities	49
Pre-Enterprise Search	49
Plan for a Staged Implementation	50
Creating the RFI or RFP	50
Collecting Your Content	51
Identify Your Content and Repositories	51
Understanding Your Users and Their Needs	51
The Acquisition Process	52

Meeting with Suppliers	52
The Proof of Concept	53
Negotiating the Purchase	54
Gathering Information: Training and References	54
Planning Your Deployment	54
Optimizing the Roll-Out	55
Tuning Relevance	55
After the Roll-Out	57
Search Sucks!	57
Theory and Practice	58
Summary	58
CHAPTER 3: OVERVIEW OF MICROSOFT ENTERPRISE SEARCH PRODUCTS	61
Understanding the 2010 Enterprise Search Product Line	62
SharePoint Foundation 2010	64
Search Server 2010 Express	64
Search Server 2010	65
SharePoint Server 2010	65
FAST Search Server 2010 for SharePoint	66
FAST Search Server for Internet Sites	66
Summary of Enterprise Search Options	67
SharePoint Server Search versus FAST Search for SharePoint	69
Summary of Feature Differences Between SP and FS4SP	70
Capabilities Found Only in SharePoint Server 2010 Search	73
Typical Search Scenarios	73
Intranet Search	74
People Search	74
Site Search	75
Research Portals	75
Customer Management	75
Parts Search	76
Intelligence and Investigation	77
Intellectual Property Protection	77
Compliance	77
E-Commerce	77
Windows Desktop Search	78
The Role of the Desktop in Enterprise Search	79
Windows Desktop Search 4: XP and Vista	79
Windows 7 Desktop Search	80

Search in Exchange and SQL	81
SQL Server 2008	82
Exchange 2010	83
Bing	84
The Bing Exploration Approach	85
Sections of Bing	85
Voice Search: Bing and Tellme	86
Federation across Products	87
Federation and OpenSearch	87
Windows 7 and Federation	88
Choosing a Product for Your Application	89
Summary	90

PART II: IMPLEMENTING SEARCH

CHAPTER 4: SEARCH WITHIN SHAREPOINT 2010	93
Searching with SharePoint 2010	93
SharePoint 2010 Overview	94
Search Server 2010	94
SharePoint Search 2010	95
Fast Search for SharePoint 2010	96
SharePoint 2010 Search Architecture Overview	96
Key Components	97
Administration	97
Crawler	97
Query and Index	98
Deployment Planning	99
Capacity Planning	99
Distribution Considerations (Location, Location, Location . . .)	102
Information Architecture	102
Crawled Properties and Managed Properties	102
Content Types	103
Installation and Configuration	104
Single Node	105
Multiple Nodes	105
Scaling a Search Application in a Farm	107
Configuration	108
People Search	109
Reporting	110
Implementing Security	111
Capacity and Limitations	112

Search Server 2010	112
SharePoint Search 2010	112
Topologies	113
Federation	114
Search User Experience	115
People Search	116
Operations	117
Relevance Tuning	117
Summary	117
CHAPTER 5: FAST SEARCH WITHIN SHAREPOINT 2010	119
Search with FAST Search for SharePoint 2010	120
FAST Search for SharePoint 2010 Product Overview	120
Key Search Engine Components	121
Administration	121
Content Gathering	122
Content-Processing Pipeline	122
Index	122
FAST Search Authorization	123
Query Component	124
Web Analyzer	124
Installation and Configuration	124
Prerequisites	125
Installation	127
Configuration	129
Multinode	133
Capacity and Limitations	133
Topologies and Federation	134
Search Cluster Columns and Rows	138
Topology Design Considerations	139
Federation	140
Content Processing	141
Tokenization, Normalization, and Lemmatization	141
Entity Extraction	142
Query Component	142
FAST Query Language	143
Relevance Tuning	145
Managed Properties	146
Rank Profile	147
Search Center	148

Search Administration	150
Monitoring	150
PowerShell	151
Search Center	151
End User Search Experience	151
Summary	155
CHAPTER 6: CUSTOMIZING SEARCH WITH SHAREPOINT 2010	157
Search Architecture and Topologies	158
SharePoint Search Key Components	158
FAST Architecture and Topology	159
How Architecture Meets Applications	162
Developing with Enterprise Search	163
Range of Customization	163
Top Customization Scenarios	164
Search-Driven Applications	165
Search-Driven Application Example	166
Customizing the Search User Experience	167
Web Parts with FAST	168
Inter-Web-Part Communication	168
Example: Tag Cloud Web Part	169
Search Connectors and Searching LOB Systems	172
New Connector Framework Features	173
Crawling Associated Content	174
Creating Indexing Connectors	174
Model Files	175
Tooling in SPD and VS2010	175
Writing Custom Connectors	176
A Few More Tips	178
Deploying Connectors	178
FAST-Specific Indexing Connectors	178
Customizing Connectivity	179
Working with Federation	179
Custom Federation to FAST ESP	181
Additional Considerations in Federation	182
Working with the Query OM	183
Query-Side APIs and OMs	183
The Federation Object Model	184
The Query Web Service	185

The Query RSS Feed	185
The Query Object Model	185
Query Syntax	186
FAST Query Language	187
Using RANK and XRANK	188
Using SortFormula	189
Content Enhancement	190
Advanced Content Processing with FAST	190
Content Pipeline Configuration	191
Content Pipeline Extensibility	191
Multilingual Search	193
Extending Search Using the Administrative OM	193
Authentication and Security	193
Search Reports	194
Summary	194

PART III: IMPLEMENTING SEARCH WITH ESP

CHAPTER 7: INTRODUCTION TO FAST ESP	197
ESP Product Overview	198
System Architecture and Administration	199
ESP Architecture and Key Components	199
The ESP Administrative Interface	201
Installation and Setup	204
Hardware	204
Operating System and Software	207
Preparing to Install ESP	210
Doing the Install	212
Surprises During Installation	214
Install Checklist	215
Quick Start: Build and Search	216
Creating a Collection	216
Searching the Collection	220
Managing ESP	222
FAST Home	223
Search Business Center	236
ESP Admin GUI	243
Other ESP Goodies	258
Documentation by Role	262
Technical Assistance	265

CTS and IMS	265
Introducing CTS and IMS	265
Content Transformation Services	266
Interaction Management Services	268
Summary of the CTS and IMS	269
Summary	269
CHAPTER 8: CUSTOMIZATION AND DEPLOYMENT OF FAST ESP 5.X	271
<hr/>	
The Big Picture: Why It's Built This Way	272
The Core Nodes	272
Getting Data into and out of the Core	273
Inbound Documents	273
Searches and Outbound Search Results	274
QR Server Versus RTS Search	275
Scaling in General	276
Scaling Core RTS Search Nodes for Content	276
Scaling Non-Core Nodes for Content	276
Scaling Cores for Search Traffic	277
Scaling in Both Directions	277
Overview: How a Multinode Installation Works	277
Multinode Installation Prep Work	278
High-Level Checklist	278
TCP/IP Configuration Should Not Be Changed After Installation	278
Check Your Firewall against ESP Ports	279
Symmetric DNS	284
Unix-Specific Considerations	284
Input Mode for Various Admin Tasks	284
You Probably Need X-Windows	285
Minimum X-Windows Resolution	285
Unix Multinode Non-GUI Installs	287
The Two Normal Scenarios	287
Other GUI Choices for Unix Installations	287
Can You Tunnel?	288
Network Settings	289
Other Choices for Bootstrapping InstallProfile.xml	289
Using a Temporary Virtual Linux Box	289
Using a Temporary Windows Microsoft Machine!	290
Custom Install Profiles	290
Specific Unix SSH Requirements	292
Summary of SSH Setup Procedure	292

Replace Passwords with SSH Credentials and Keys	293
Removing SSH Host Authorization Prompts	295
Distributing the SSH Credentials	296
Reviewing a Three-Node Linux Installation	298
Allocation of Processes	298
Installation Profile	300
Starting the Installation	302
Defining Nodes	305
System Tests	306
Assigning Processes	307
Copying Node Files	314
Installing	317
Backing Up Quiescent States	318
A Few ESP Terms You Might Come Across	319
About Python and FAST Cobra	319
References to FDS	319
The Mythical ESPDeploy Application	319
Customizing the Document Processing Pipeline	320
About ESP's Internal Data Representations	321
Managing Pipelines Processing Stages	322
Simple Pipeline Modification Example	322
Configuring the Pipeline	323
Step 1: Create the Batch File	323
Step 2: Create a Document Pipeline Stage	324
Step 3: Create a Document Pipeline Using the New Stage	325
Step 4: Create a Collection and Submit a Single Document	326
Step 5: Verify That Your Stage Works	326
Some Caveats	326
ESP Template Pipelines and Stages	327
Summary	332
CHAPTER 9: ADVANCED TOPICS	333
<hr/>	
Scalability	334
ESP Rows and Columns	334
Scaling the Amount of Data	334
Latency and Updates	335
Metadata and Pipelines	335
Licensing and Content	335
Scaling the Number of Searches	335
Average and Peak QPS	335
Failover Rows	336

Don't Bet Entirely on Caching	336
Integrating with a Taxonomy	336
Background and Definitions	336
Types of Taxonomies	337
Taxonomies Versus Ontologies	339
Other Terms Associated with Taxonomies	339
Faceted Navigation and Taxonomies	340
Taxonomies as a Symptom?	340
Taxonomy Workflow	341
Taxonomy Rule Sets	341
FAST Partner Teragram	342
Rolling Your Own	342
A Brief Overview of Other Advanced ESP Features	348
Structural Analysis Toolkit	348
How Entities and Fields Relate to Navigators, Refiners, and Facets	348
Various Methods for Entity Extraction	351
STAN Workflow	352
Relevancy Tweaking	353
Default Relevancy	353
Relevancy Tuning Options in ESP	354
FAST's Hybrid Security Model	355
Unsupervised Clustering and Noun Phrase Extraction	357
Federated Search	358
Summary	361

PART IV: ENTERPRISE SEARCH 2.0

CHAPTER 10: ENTERPRISE SEARCH IS SOCIAL SEARCH	365
Social Computing	366
Social Search	366
Types of Social Activities That Affect Search	368
Social Search in SharePoint 2010	370
Outlook 2010	370
Implementing Social Search in SharePoint 2010	371
User Profile Service Application	371
User Profile Synchronization	372
My Site	373
User Profiles	375
People Search	378
Refiners	379
Self-Search	380

Social Search Experience	381
Suggestive Search	381
Contextual Search	382
Click-Through Relevancy	383
Social Tags	384
Implicit User Profile Data	385
Extending Social Search	386
Enhancing Search Through Content-Processing	386
Leveraging Social Distance Information	386
Content Rating and Reviews	386
Summary	387
CHAPTER 11: SEARCH AND BUSINESS INTELLIGENCE	389
<hr/>	
Search and BI Are Converging	390
Search and BI Combine in Different Ways	391
Text Analytics	392
Search Analytics	394
Unified Information Access	397
Business Intelligence Needs Search	399
The State of Traditional BI	399
Challenges with Traditional BI	400
Benefits of BI Search Combinations	401
BI for the Masses	401
Search as the UI Paradigm for BI	402
Adoption of Search/BI Integration	403
Survey on Adoption	403
Areas of Adoption	404
Applications at the Intersection of Search and BI	405
Customer-Centric Applications	405
Information Consolidation	406
Operational Intelligence	406
Microsoft BI Overview	406
BI in Microsoft SharePoint 2010	406
PowerPivot for Excel 2010	409
SQL Server	410
SSRS	410
SSRS with SharePoint	411
Microsoft StreamInsight and Complex Event Processing	411
FAST for Search on Structured Data	413
The Microsoft BI Indexing Connector	415
Tour of the Reports Center	415

Features of the BI Indexing Connector	416
Architecture of the BI Indexing Connector	417
Setup and Administration of the BI Indexing Connector	418
The Future of Search and BI	421
Summary	421
CHAPTER 12: THE FUTURE OF SEARCH	423
<hr/>	
The Sorry State of Search	423
The Flawed HAL9000 Usage Model	423
Enterprise Search 2.0	424
Business-Driven Search	425
Conversational Search	425
Socially Aware	425
Integrated with Other Content Sources	426
Emerging Trends in Enterprise Search	426
Business-Based Goals	427
User Experience	428
Search Technology	432
Operations	434
Enterprise Search 3.0 — and Beyond	436
What You Can Do Today	438
Summary	439
APPENDIX: RESOURCES	441
<hr/>	
Books	441
Microsoft Enterprise Search Sites	442
User Forums	442
Newsletters	442
Blogs	443
Vendor Blogs	443
Other Blogs	444
Trade Shows	444
INDEX	445