

ENTERPRISE CONTENT AND SEARCH MANAGEMENT

FOR BUILDING DIGITAL PLATFORMS

Shailesh Kumar Shivakumar




IEEE PRESS

 IEEE
computer
society

WILEY

Enterprise Content and Search Management for Building Digital Platforms



IEEE Press Editorial Board

Tariq Samad, *Editor in Chief*

George W. Arnold	Xiaoou Li	Ray Perez
Giancarlo Fortino	Vladimir Lumelsky	Linda Shafer
Dmitry Goldgof	Pui-In Mak	Zidong Wang
Ekram Hossain	Jeffrey Nanzer	MengChu Zhou

Kenneth Moore, *Director of IEEE Book and Information Services (BIS)*

About IEEE Computer Society

IEEE Computer Society is the world's leading computing membership organization and the trusted information and career-development source for a global workforce of technology leaders including: professors, researchers, software engineers, IT professionals, employers, and students. The unmatched source for technology information, inspiration, and collaboration, the IEEE Computer Society is the source that computing professionals trust to provide high-quality, state-of-the-art information on an on-demand basis. The Computer Society provides a wide range of forums for top minds to come together, including technical conferences, publications, and a comprehensive digital library, unique training webinars, professional training, and the TechLeader Training Partner Program to help organizations increase their staff's technical knowledge and expertise, as well as the personalized information tool myComputer. To find out more about the community for technology leaders, visit <http://www.computer.org>.

IEEE/Wiley Partnership

The IEEE Computer Society and Wiley partnership allows the CS Press authored book program to produce a number of exciting new titles in areas of computer science, computing, and networking with a special focus on software engineering. IEEE Computer Society members continue to receive a 15% discount on these titles when purchased through Wiley or at wiley.com/ieeecs.

To submit questions about the program or send proposals, please contact Mary Hatcher, Editor, Wiley-IEEE Press: Email: mhatcher@wiley.com, Telephone: 201-748-6903, John Wiley & Sons, Inc., 111 River Street, MS 8-01, Hoboken, NJ 07030-5774.

Enterprise Content and Search Management for Building Digital Platforms

Shailesh Kumar Shivakumar

IEEE
computer
society


IEEE PRESS

WILEY

Copyright © 2017 by the IEEE Computer Society, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Names: Shivakumar, Shailesh Kumar, author.

Title: Enterprise content and search management for building digital platforms / Shailesh Shivakumar.

Description: Hoboken : Wiley, 2016. | Includes index.

Identifiers: LCCN 2016032953 (print) | LCCN 2016048578 (ebook) |

ISBN 9781119206811 (paperback) | ISBN 9781119206828 (pdf) |

ISBN 9781119206835 (epub)

Subjects: LCSH: Management—Technological innovations. | Digital media—Management. |

Multimedia systems—Management. | Performance technology. |

BISAC: COMPUTERS / Web / Site Design.

Classification: LCC HD30.2 .S558 2016 (print) | LCC HD30.2 (ebook) |

DDC 658.4/038011—dc23

LC record available at <https://lccn.loc.gov/2016032953>

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

*To my parents, Shivakumara Setty V and Anasuya T M,
from whom I borrowed love and strength
To my wife, Chaitra Prabhudeva, and my son, Shishir,
from whom I borrowed time and support
To my in-laws, Prabhudeva T M and Krishnaveni B,
from whom I borrowed help and courage
and
To all my schoolteachers who bestowed lots of love and
knowledge upon me*

Contents

Preface	xvii
Acknowledgments	xxvii
About the Author	xxix
About the Companion Website	xxxii

Part 1 Content Management Basics for Digital Platforms

1 Introduction to Digital Platforms	3
1.1 Enterprise Digital Ecosystem	4
Digital Opportunities for Enterprises	4
Challenges in Modern Digital Enterprises	5
Enterprise Digital Capabilities	6
Digital Disruption across Business Domains	9
1.2 Concepts of Enterprise Content Management (ECM)	15
Enterprise Content Ecosystem	15
Content Presentation	16
Content Applications	16
Enterprise Services	18
Access Channels	18
ECM vs WCM	19
Book's Focus Areas	19
1.3 Enterprise Digital Strategy and Content Strategy	20
Enterprise Digital Strategy	21
Core Digital Technologies	24
Digital Content Strategy	25
1.4 Digital Content Management and Enterprise Search:	
An Overview	28
Digital Content Management	28
Enterprise Search	29
Enterprise Sources Used by Search Engine	30
1.5 Chapter Summary	30
2 Content Strategy	32
2.1 Overview of Content Strategy	32
Introduction to Content Strategy	33

	Main Tenets of Content Strategy	33
	Related Strategies and Artifacts Used in Content Strategy	34
	Common Challenges with Content	36
	Definition of Key Terms	38
2.2	Prerequisites for Content Strategy	38
	Content Requirements	39
2.3	Defining Content Strategy	41
	Phases of Content Strategy	41
	Design Considerations	46
	Core Elements of Content Strategy	50
2.4	Content Strategy Case Study	73
2.5	Chapter Summary	79

3 Basics of Content Management Systems

82

3.1	What Is a Content Management System?	82
	Business Drivers for CMS	84
	Utilities of CMS	85
	Challenges in Implementation of CMS	86
	CMS Trends	87
	Various Roles in CMS	89
3.2	CMS Key Design Principles	89
	Reusability and Flexibility	89
	Taxonomy and Metadata Definition	90
	Standards Definition	90
	Consolidated Content Repository	91
	Governance Model	91
	Content Syndication and Services	91
	CMS Evaluation	91
3.3	CMS Capabilities and Attributes	92
	Desired Core Capabilities of CMS	92
	Main Functionality of Core CMS components	94
	Desired Attributes of CMS	96
3.4	Content Lifecycle Management in CMS	98
3.5	A Brief Description of Open Source CMS and JCR	100
	Drupal (https://www.drupal.org/)	100
	Joomla (https://www.joomla.org/)	101
	WordPress (https://wordpress.org/)	101
	JCR Implementation: Apache Jackrabbit	102
3.6	Chapter Summary	102

4 Content Management System Architecture

104

4.1	CMS Design and Architecture	104
	CMS Implementation Approach	105
4.2	Modern CMS Architecture Patterns	106
	MVC Architecture	106
	N-tier Architecture and N-layer Architecture	108

	Service-Oriented Architecture (SOA)	109
	Microservices Architecture	111
4.3	CMS Value Articulation and Solution Principles	113
	CMS Value Articulation Framework	113
4.4	CMS Solution Design Principles	114
4.5	Design of CMS Solution Components	118
	Multi-Site Management Design	118
	Content Folder Design	124
	Content URL Design	124
	Localization Design	126
	Collaboration Design	129
4.6	CMS Operations Management	130
	Release Management Process	131
	CMS Maintenance	134
4.7	Realizing Content Strategy with CMS	137
	Content Design and Creation Using CMS	137
4.8	CMS Reference Architectures	137
	Customer Experience Platform (CXP) Reference Architecture	137
	Knowledge Management System Based on CMS	143
	Digital Marketing Platform Based on CMS	147
	Architecture of Apache Jackrabbit	147
4.9	Chapter Summary	152

5 Development Using Templates and Workflows

154

5.1	CMS Template Design	154
	What Are Templates?	155
	Authoring Template, Presentation Template, and Page Layout	155
	Design of Authoring Templates	157
	Content Templates Based on Open Standards	158
	Content Presentation Templates	159
5.2	Authoring Content Using an Authoring Template	160
	Template-User Interface	162
	Using Templates for Pages	162
5.3	Chunking and Templates for Chunks	165
	Design Considerations for a Content Chunk	168
	Content Chunk-Based Page Content Aggregation	168
	Case Study: Chunk Identification and Chunk Template Design for Product Pages	170
	Template Guidelines and Best Practices	176
5.4	Template Support among Various CMS	178
5.5	Case Study: Building Content Templates for a Web Support Site	179
	Web Support Site Case Study: Context and Overview	179
5.6	Content Workflows	183
	Workflow Design	184
	Designing Workflow	187
	Workflow Optimization	188

5.7 Case Study: Modeling Workflow for a Knowledge Management System 189
5.8 Chapter Summary 192

6 Content Information Architecture, Taxonomy, and Metadata 195

6.1 Intuitive Information Architecture 196
Goals of IA 196
Elements of IA 196
Defining IA 196
Best Practices While Defining IA 198
Role of IA in Content Strategy 199
IA Design Samples 200
6.2 Introduction to Taxonomy and Metadata 202
Advantages of Taxonomy and Metadata 203
Business Drivers for Taxonomy and Metadata 204
Taxonomy and Metadata Best Practices 204
Types of Metadata 205
Metadata Hierarchy Modeling 205
6.3 Metadata Usage in Relevant Content Discovery 208
6.4 Integration of Metadata with CMS 208
6.5 Metadata Standards and Formats 210
Dublin Core 210
Simple Knowledge Organization System 211
6.6 Case Study: Content Metadata to Increase Search Effectiveness 212
Internal Search 213
External Search 214
6.7 Other Utilities of Content Metadata 214
Metadata-Based Content Categorization 214
Marketing and Sales Support 214
Metadata-Driven Content Personalization 214
Metadata-Based Page Customization 215
Content Metadata for Navigation 215
Analytics Metadata 215
Content Metadata for Workflow 215
Using Metadata for Reusing Content Chunks 215
Security Metadata 216
6.8 Taxonomy Governance 216
Social Tagging 217
6.9 Chapter Summary 217

Part 2 Advanced Content Management

7 Content Integration and Content Standards 221

7.1 Content Integration Requirements 221

7.2	CMS Integration View	222
	Enterprise CMS: The Big Picture	223
7.3	CMS Integrations	225
	Security Integration	225
	Translation System Integration	226
	Search Engine Integration	228
	Content Services Integration	229
	Portal Integration	231
	Presentation Engine Integration	232
	Metadata Management System (MMS) Integration	232
	Feed Integration	233
	Digital Asset Management (DAM) Integration	233
	JCR-Based Integration	234
7.4	CMIS-Based Integration	235
	When can we use CMIS-based integration?	236
	JCR and CMIS	236
7.5	CMS Integration with Other Systems	237
7.6	Content Standards	237
	HTML/XHTML	239
	XML (Extensible Markup Language)	239
	DITA (Darwin Information Typing Architecture)	242
	JSON (JavaScript Object Notation)	244
	SCORM (Sharable Content Object Reference Model)	244
	Feed Formats: RSS/ATOM	246
	Web Service Standards: SOAP and REST	247
7.7	Chapter Summary	250
8	Digital Asset Management and Document Management	253
8.1	Digital Asset Management (DAM)	254
	DAM Definition	254
	DAM Objectives and Trends	254
	Need for DAM	255
	DAM Business Scenarios	256
	Architecting an Enterprise DAM System	257
	DAM Challenges and Best Practices	261
8.2	Document Management	263
	Capabilities of Document Management System	263
	Document Management Elements and Functions	265
	Document Management Evolution and Road Map	266
	Case Study: Document Management Solution for a Banking Portal	266
8.3	Chapter Summary	270
9	Content Migration	272
9.1	Content Migration	272
	Content Migration Drivers	272
	Content Migration Principles	273

xii Contents

Migration Design Considerations	275
Migration Challenges and Best Practices	276
Migration Checklist	279
Migration Approach	279
Content Migration Examples	284
Migration Governance	287
Migration Automation	288
Cutover Plan	290
Migration Case Study: JCR-Based Custom Migration Script	291
9.2 Chapter Summary	295

10 Content Governance: Validation, Analytics, KPIs, SEO, and Evaluation

297

10.1 Content Validation	298
Content-Testing Checklist	300
10.2 Content Analytics and KPIs	304
Content Analytics	304
Content KPIs	310
10.3 Content SEO	312
Content SEO Strategy	312
Content SEO Best Practices	313
10.4 CMS Evaluation Framework	315
Business Considerations for Selecting a CMS Product	315
Evaluation Framework	317
10.5 Appendix: WCMS Features	322
10.6 Chapter Summary	325

11 Content Security

327

11.1 Content Security Vulnerabilities and Mitigation Steps	327
Cross-Site Scripting (XSS)	328
SQL Injection Attacks	329
Denial of Service (DoS) and Distributed Denial of Service (DDoS)	330
Cross-Site Request Forgery (CSRF)	331
Clickjacking	332
11.2 Generic Content Security Scenarios	333
Authentication and Authorization	333
Single Sign-on (SSO)	334
Permission Model Using Roles and Permissions	334
11.3 Security Testing	337
Core Security Testing	337
Security Code Reviews	337
Penetration Testing	338
11.4 Security Best Practices	339
Adopt Multi-Layer Security	339
Robust Account Management	340

	Proactive Scanning and Vulnerability Assessment	340
	CMS Patching and Upgrades	340
	Transport-Level Security	341
	CMS Hardening	341
	Security Logging and Auditing	342
	Content Archival and Backup	342
	Content Classification	342
	Disaster Recovery and Business Continuity	342
	Restricted File Permissions	343
	Security Monitoring	343
	Iterative Security Testing	343
	Error Handling and Resource Handling	343
	Security Governance	344
	Web Application Firewall (WAF) and Security Plugins	344
	Hosted Content Systems	344
11.5	Case Study: Security Testing for a CMS Application	344
	Application Background	344
	Security Testing Details	345
	Security Vulnerabilities and Remediation Measures	345
	Application Retesting	350
11.6	Chapter Summary	350

12 Content Infrastructure and Performance Optimization **352**

12.1	CMS Infrastructure Architecture	352
	Infrastructure Sizing	353
	Basic Concepts of CMS Deployment Architecture	354
	CMS Deployment Setup	355
	Disaster Recovery Setup	357
12.2	Content Performance Optimization	358
	Optimal Content Design	358
	Optimal Page Design	358
	Optimized Publishing Workflows	360
	Database-Level Performance Optimizations	360
	Content Caching Design	360
	Monitoring and Notification Setup	360
	Proactive Identification of Memory Leaks	361
	CMS-Level Performance Optimizations	361
	Logging	362
	CMS Caching	362
	Search-Engine-Level Performance Optimization	362
	Infrastructure-Level Performance Optimization	362
12.3	Content Performance Key Performance Indicators (KPIs)	364
	Collecting Content Performance KPIs	364
	Content Performance KPIs	364
12.4	Content Performance Validation	365

12.5	Content-Related Best Practices	366
	Content Best Practices	366
	Content Best Practices Checklist	367
	CMS Best Practices	370
12.6	Chapter Summary	373

Part 3 Enterprise Search Technologies

13 Introduction to Enterprise Search **377**

13.1	Introduction to Enterprise Search	378
	Business Drivers	378
	Technology Drivers	380
	Challenges of Enterprise Search	381
	Generic Best Practices of Enterprise Search	382
13.2	Enterprise Search Overview	383
	Enterprise Search Architecture Layers	383
	Web Search vs. Enterprise Search	384
	Search-Related Trends	386
	Search Evolution	388
	Key Value Proposition of Enterprise Search	389
13.3	Enterprise Search capabilities	389
13.4	Enterprise Search Features	392
	Basic Search Features	392
	Advanced Search Features	393
	Features in Apache Solr and ElasticSearch	396
13.5	Chapter Summary	397

14 Advanced Enterprise Search **398**

14.1	Federated Search	398
	Features of Federated Search	399
	Sample Federated Search Architecture	399
	Common Challenges with Federated Search	400
	Enterprise Search through Intermediate Aggregation Repository (Alternative to Federated Search)	402
14.2	Advanced Search Features	403
	Relevancy Rank Adjustment and Rank Boosting	403
	Personalized Search	404
	Alternative Search Suggestion	405
	Secured Search	405
14.3	Enterprise Semantic Search	409
	Key Elements of Semantic Search	410
	Enterprise Semantic Search Architecture	411
	Enterprise Semantic Search Process	411
	Semantic Search Capabilities in Apache Solr	412

14.4	People Search and Social Search	412
	Challenges in People Search	413
	People Search Design	413
	Sample People Search in Apache Solr	414
	Social Search	414
14.5	Mobile Search	415
14.6	Big Data Search	415
	Apache Solr Integration with Apache Hadoop	416
14.7	Search Engine Optimization (SEO)	417
	Page-Level SEO Tags	417
	SEO Strategy	419
	SEO Best Practices	420
	SEO Anti-Patterns	421
14.8	Case Study: Information Management Portal Driven by Apache Solr	422
	Background and Context	422
	Information Management Portal Solution Components	422
14.9	Chapter Summary	424

Further Reading 427

Index 429

Preface

Disruption in digital technologies has opened up an entirely new realm of possibilities for enterprises. Harvesting new-age digital technologies can redefine the ways business is done online and can potentially give numerous possibilities to reengage with stakeholders such as consumers, partners, resellers, and others. Digital technologies enable enterprises to provide on-demand, customer-centric, personalized, contextual, and meaningful content from anywhere, anytime, on any device. Digital-enabled business models reshape customer experiences and form the key differentiators. As a result, the digital user will be meaningfully engaged bringing in productivity, loyalty, and long-term relationship. On the B2B front, digital technologies have also opened up new realms of possibilities through process optimizations, enterprise integrations, and other developments, and a digital technology ecosystem has reshaped infrastructure and operations sides of things through hardware consolidation and cloud enablement.

Digital technologies are disrupting most of the business domains, technology ecosystems, and business processes. Due to its wide range of benefits and long-term strategic impact and competitive advantages, enterprises across domains are embracing digital revolution. In today's hyper-connected world, word-of-mouth promotion is given preference over sponsored ads, Facebook "likes" count more than expert rating, and enterprises strive hard to convert a visitor into a brand advocate using digital technologies.

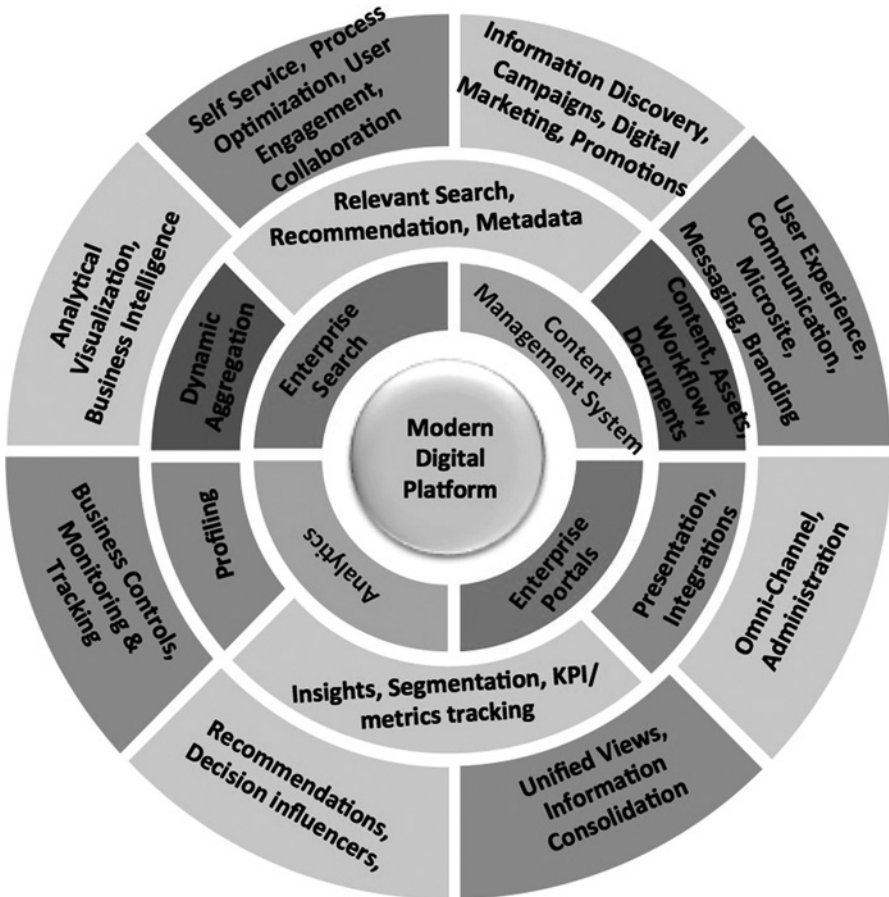
DRIVERS AND MOTIVATIONS FOR THE BOOK

Modern enterprises face multiple challenges in building a robust enterprise digital ecosystem. The challenges are multifold in nature and consists, among other things, of internal challenges concerning employee productivity, process optimization, information management, content management, and big data management. Coupled with these are external challenges such as, among others, Omni-channel customer engagement, social and collaboration integrations, personalized presentation, and competitive pressures.

Based on our experience, the most effective way to address these challenges is to provide a *robust information management system* consisting of *seamless relevant information discovery*. This book tries to address these two fronts by exploring various concepts in digital content management (for information management) and enterprise search (for efficient information discovery). This book takes a differentiated view

through a combined focus on content management and enterprise search. During the process it aims to help organization build robust digital platforms using proven best practices, practical models, and time-tested techniques discussed in the book.

We can map the technology topics (CMS and Enterprise Search) discussed in this book with the modern digital platform's capability as shown in the following diagram:



The first layer depicts the core digital technologies, namely content management system (CMS), enterprise search, portals, and analytics. The second layer maps the technical capabilities offered by corresponding technologies. CMS provides robust content management, workflows, documents, and asset management whereas search provides relevant search and recommendations. Both CMS and search would enable metadata-tagging capabilities. The outermost layer depicts the business capabilities

enabled by corresponding technology capabilities. Content management enables intuitive user experience, communication, messaging, branding, and micro-site. Search and CMS combined enables promotions management, campaigns, and marketing and relevant information discovery capabilities.

The diagram depicts the importance of role played by CMS and enterprise search in a digital scenario. CMS and search form the information management backbone for a digital enterprise. The book tries to cover the capabilities discussed under CMS and search umbrella and relevant analytics capabilities wherever applicable.

Key Differentiators of the Book

The key differentiators and novel aspects of this book are summarized in the following list:

- **Wide coverage of modern methodologies and techniques:** We have covered emerging technologies such as micro services architecture in content management, CMS-based customer experience platform (CXP), Big Data search, semantic search, Omni-channel content enablement, JCR and CMIS standards, content analytics, SEO, and KPIs. We have detailed trends in CMS and enterprise search we have noticed and have provided good coverage of emerging trends. CMS is explored from security, infrastructure, and performance viewpoint as well.
- **Content frameworks:** The book covers many practically proven models and techniques related to CMS evaluation framework, content migration framework, search evaluation framework, and other aspects that can be used in real-world digital engagements. Comprehensive CMS, search, and DAM evaluation templates are given in Appendixes C, D, and F, respectively.
- **Elaborate content strategy discussion:** As content strategy forms the core of content management, we engage in an in-depth discussion of it in Chapter 2 along with a detailed case study. All chapters in Parts I and II are organized to realize various elements of content strategy discussed in Chapter 2. We have also provided a content strategy template in Appendix A to complement the concepts discussed in Chapter 2.
- **Case-study-based approach:** All core topics (such as templates, workflows, content security, performance, metadata, document management, content migration, and such) have detailed in-context case studies to provide the practical flavor to the topic discussion. The Online Wiley book support material section provides content case studies to explain the best practices used in real-world engagement. Case studies are used as tools to reinforce the theory concepts and provide practical applicability for them. Online support material

also has an elaborate end-to-end digital program case study covering CMS and enterprise search for a digital e-commerce platform.

- **Sample code and configuration:** We have provided sample code while discussing JCR migration concepts to elaborate the concept in Chapter 9. We have also given the configurations that can be used to address security vulnerabilities and optimize content performance in Chapters 11 and 12, respectively.
- **Reference architectures:** The book provides reference architecture for various CMS and search-based applications. Reference architecture of CMS-based customer experience platform, knowledge management system, digital marketing platform, and e-commerce platform are elaborated.
- **Proven best practices and checklists:** We have provided elaborate practically proven best practices while discussing key topics (such as content services, content security, templates, etc.). We also provided content management checklist in Appendix B section. Architects and managers for content and search engagements could use this.
- **Content integrations:** We have dedicated Chapter 7 to integrations with CMS providing details about optimal integration techniques with CMS.
- **Synergies between enterprise content and search:** This book tries to explore the synergies between enterprise content and search systems to build a robust digital platform. Metadata, taxonomy, SEO, analytics, and digital program management are explored from this dimension.
- **Practically proven models and best practices:** We discuss various models and best practices related to content such as template design, workflow design, and Omni-channel content design that are successfully employed in various practical engagements.
- **Architecture concepts:** There is an in-depth coverage of various architecture concepts for content management and digital search. Practitioners can use this as reference architecture in digital programs.
- **Reusable templates:** We have provided CMS evaluation template, search product evaluation, and content strategy template in the appendix sections. Readers can use it for content programs.

MAIN THEMES AND FOCUS AREAS

Main themes and focus areas of this book are:

- **Digital content management and enterprise search:** The primary focus areas of this book are digital content management (primarily Web content and digital assets through Web content management [WCM] concepts) and enterprise search. Wherever necessary, the book also elaborates other supporting

systems/components such as digital asset management (DAM) systems, document management system, workflow management, and Web analytics, among others.

- **Technology and product agnostic view:** The concepts, methodologies, techniques, and best practices discussed in the book are product and technology agnostic. Wherever necessary, concrete examples are drawn from specific technologies and products to explain the concept.
- **Open source frameworks:** Many of the concrete examples are drawn from open source products. Some reference architectures are also developed using open-source components. The intention is to help readers leverage open-source technologies while creating digital systems.
- **Proven practical methodologies and best practices:** We have elaborated many proven models and best practices in areas such as content migration, CMS evaluation framework, content performance, content security, and such. This would help the content and search practitioners apply these frameworks and techniques.
- **Challenges and best practices:** While discussing core portal technologies such as integrations, content management, search, and others, we have discussed the commonly encountered challenges/pitfalls and the best practices.

CHAPTER ORGANIZATION AND TARGET AUDIENCE

The book is organized in three parts with 14 chapters. The online Wiley book support material section provides various supporting material such as content case study and end-to-end digital case study. Part I consists of six chapters that introduce reader to core concepts of content management. We look at content strategy, CMS basics, CMS architecture, templates and workflow, information architecture, taxonomy, and content metadata. Part II includes six chapters and extends the content management concepts and elaborates on topics related to integration, content standards, DAM and document management, content migration, CMS evaluation, content validation, content analytics, content security, content performance. Part III consists of two chapters and is mainly dedicated to discussing basics of enterprise search and advanced search.

We have provided six appendix sections: Appendix A provides a content strategy template, Appendix B provides a checklist for various content management activities, Appendix C is a CMS product evaluation template, Appendix D is the enterprise search product evaluation template, Appendix E provides sample Java code for adding a JCR node, and Appendix F provides an evaluation template for DAM platforms.

The following is the high-level summary of various chapters along with intended target audience:

Chapter	Main topics	Target Audience
Part I: Content Management Basics For Digital Platforms		
Chapter 1: Introduction to Digital Platforms	Enterprise digital ecosystem, enterprise content management concepts, digital strategy, content strategy, digital content management overview, enterprise search overview	Digital architects, enterprise architects, program managers, business analysts, and senior business executive
Chapter 2: Content Strategy	Content strategy overview, strategy challenges, strategy essentials, content characteristics, requirements elaboration, content strategy definition process, content strategy phases, content strategy elements, content strategy case study	Content architects, content strategists, CMS developers, and enterprise architects
Chapter 3: Basics of Content Management System	CMS drivers, CMS design principles, CMS attributes, CMS capabilities, Discussion of CMS systems (WordPress, Drupal, Joomla), and Apache Jackrabbit	Enterprise architects, content architects, and CMS developers
Chapter 4: Content Management System Architecture	CMS design and architecture, CMS architecture patterns, CMS value articulation framework, CMS solution component design, Multi-site design, content folder design, content URL design, localization design, CMS infrastructure design, content strategy realization in CMS, CMS reference architecture, customer experience platform design, knowledge management system design, and content marketing platform design	CMS architects, CMS developers and enterprise architects
Chapter 5: Templates and Workflows	CMS template design, authoring and presentation templates design and usage, template-user interface, template development case study, workflow design, workflow optimization, workflow case study	Content architects, content authors, and CMS developers
Chapter 6: Content Information Architecture, Taxonomy, and Metadata	Designing intuitive information architecture (IA), elements and goals of IA, taxonomy concepts, taxonomy and metadata, metadata types, metadata standards (Dublin Core and SKOS), metadata case study	Enterprise architects, content architects, content strategists, and information architects

Chapter	Main topics	Target Audience
Part II: Advanced Content Management		
Chapter 7: Content Integration and Content Standards	<p>CMS integrations with security systems, TMS, search engine, portals, presentation engines, metadata systems, feeds, DAM, CMIS integrations.</p> <p>Content standards: HTML/XML, DITA, JSON, SCORM, Web service formats (REST and SOAP)</p>	Content architects, Integration architects, CMS administrators, CMS developers, and enterprise architects
Chapter 8: Digital Asset Management and Document Management	DAM objects, architecting DAM system, DAM challenges, document management system capabilities, document management evolution and road map, document management case study	Content architects, CMS developers, content authors
Chapter 9: Content Migration	Migration drivers, migration design considerations, migration challenges, migration checklist, migration governance, migration automation, migration case study	Enterprise architects, CMS architects, CMS developers and program managers
Chapter 10: Content Maintenance – Validation, Analytics, KPI, SEO, and Evaluation	Content validation types, validation checklist, content analytics, content KPIs, content analytics design, content analytics case study, content SEO strategy, content SEO best practice, CMS evaluation framework	CMS architects, CMS developers, CMS QA team and program managers
Chapter 11: Content Security	Content security vulnerabilities, XSS, CSRF, denial of service, clickjacking, generic content security scenarios, SSO, penetration testing, security best practices, security testing case study	Content architects, security architects, enterprise architects
Chapter 12: Content Infrastructure and Performance Optimization	Content performance optimization, CMS-level performance optimization, infrastructure-level performance optimization, content performance KPIs, content performance testing	Content architects, enterprise architects, performance engineers, CMS developers

Chapter	Main topics	Target Audience
Part III: Enterprise Search Technologies		
Chapter 13: Introduction to Enterprise Search	Enterprise search drivers, search overview, search trends, search evolution, search capabilities, basic search features, advanced search features, Apache Solr and Elasticsearch features	Enterprise search architects, information architects, enterprise architects, and search developers
Chapter 14: Advanced Enterprise Search	Federated search, features, architecture and challenges of federated search, relevancy rank adjustment, personalized search, semantic search, semantic search process, people search, Big Data search	Search architects, enterprise architects, program managers, and search developers

DECLARATION

- Utmost care has been taken to ensure the accuracy and uniqueness of the book content. Any inaccuracies or inconsistencies are entirely my own. If you think any corrections are needed, or for any other feedback, please write to Shailesh.shivakumar@gmail.com
- In a few chapters I have used the features of popular and open-source WCM products to explain the concepts. The explanation is for educational purposes only and should not be considered as a product or technology recommendation or evaluation. The CMS plugins and modules used to illustrate examples and concepts are also for educational purposes only; they should not be interpreted as recommendations or evaluations. Comprehensive evaluation template and framework are provided in the appendix section.
- All open-source tools mentioned are in public domain as open source at the time of writing of this book.
- I acknowledge the trademarks of all products, technologies, and frameworks being used in this book.
 - WordPress, Joomla, Drupal, and dotCMS are registered trademarks and are the legal property of their respective owners.
 - Documentum is a registered trademark of EMC Corporation.
 - Oracle, Oracle Access Manager, WebCenter, WebLogic, OHS, and Java are registered trademarks of Oracle and/or its affiliates.
 - Synaptica is the registered trademark of Synaptica, LLC.
 - SiteMinder is the registered trademark of CA Technologies.
 - WebSphere, Tivoli Access Manager, IBM WCM, IHS, and DB2 are registered trademarks of IBM and/or its affiliates.
 - AEM and CQ5 are registered trademarks of Adobe and/or its affiliates.

Index

- adaptive content, 62–65, 78–79
 - case study, 66
 - developing adaptive content, 64
 - prerequisites, 63
 - support in various CMS, 65
- Adaptive Web Design, 65–66
- Apache Jackrabbit, 102, 237, 249
- ATOM, 91
- authoring template, 155–158
 - design considerations, 155
- AWD, *see* Adaptive Web Design

- best practices, 361
 - CMS best practices, 370–373
 - content best practices, 366–367
- Big Data search, 415–416

- clickjacking, 332
- CMIS, *see* content management interoperability services
- CMS, *see* content management system
- CMS architecture, 104, 106
- CMS caching, 362
- CMS deployment process, 132
- CMS evaluation framework, 315
 - business considerations, 315
- CMS implementation, 105
 - approach, 105–106
- CMS maintenance, 134–135
 - CMS backup, 134
 - monitoring, 135
 - patches and upgrades, 135
- CMS reference architecture, 137–147
 - CXM architecture, 137
 - digital marketing, 147
 - knowledge management system, 143
 - CMS solution, 118–124
 - content folder design, 124
 - content URL design, 124
 - design principles, 114
 - localization, 127
 - multi-site authoring, 121
 - multi-site deployment strategy, 123
 - multi-site management design, 118
- CMS value articulation framework, 113–114
- content analytics, 304
 - case study, 309
 - design and implementation, 305
 - impact on content strategy, 307
- content archival, 95
- content caching, 360
- content challenges, 36
- content chunk, 65–171, 210, 213, 239, 242
 - case study, 170
 - chunk template development, 170
 - chunk template identification, 170
 - content chunk management, 171
 - design considerations, 168
 - page content aggregation, 169
- content governance, 46
- content integration, 221–233
 - API integration, 232
 - content services, 229
 - feed integration, 233
 - metadata management system integration, 232
 - portal integration, 231
 - requirements, 221
 - search engine integration, 228
 - translation system integration, 222
- content KPIs, 310
- content localization, 128

- content management interoperability
 - services, 235
 - JCR and CMIS, 236
- content management system, 82–96
 - business drivers, 84
 - capabilities and attributes, 92
 - capability model, 94
 - challenges, 86
 - CMS roles, 89
 - content services, 93
 - design principles, 89
 - desired attributes, 96
 - trends, 88
 - utilities, 85
- content mapping, 53
- content migration, 272
 - approach, 277, 279
 - automation, 288
 - challenges and best practices, 276
 - checklist, 279
 - design considerations, 275
 - drivers, 272
 - extraction, 283
 - load, 289
 - migration governance, 274–275
 - principles, 273
 - transformation, 283
- content models, 53
- content performance, 358
 - infrastructure level, 362
 - KPIs, 364
 - search engine level optimization, 362
 - validation, 365
- content publisher, 96
- Content Repository API for Java, 90, 100, 116, 234, 249
- content requirements, 39
 - non-functional requirements, 40
- content reusability, 59, 157, 165, 197, 242
- content security, 327–349
 - case study, 344
 - vulnerabilities, 327
- content SEO, 312
 - best practices, 313
 - strategy, 312
- content standards, 237
- content strategy, 21, 25, 33, 69, 155, 186, 199
 - case study, 73
 - content governance, 28
 - content lifecycle, 27
 - content roadmap, 76
 - defining content strategy, 41
 - execution, 48
 - maintenance, 42, 49
 - overview, 32
 - phases, 41
 - planning, 42
 - pre-requisites, 38
 - process, 28
 - scope items, 27
- content-testing checklist, 300
- content validation, 298
 - endurance testing, 300, 304
 - functionality testing, 298
 - integration testing, 299
 - load testing, 299
 - localization testing, 299
 - performance testing, 299
 - security testing, 303
 - stress testing, 300
 - usability testing, 299
- content workflows, 183
- cross-site request forgery, 331
 - prevention, 332
- cross-site scripting, 328
 - prevention, 328
- CSRF, *see* cross-site request forgery
- customer experience management, 137
- cutover plan, 277, 290
- CXM, *see* customer experience management

- DAM, *see* digital asset management
- Darwin Information Typing Architecture, 244
- DDoS, *see* distributed denial of service
- delta migration, 291
- denial of service, 330
- deployment model, 131
- digital asset management, 233, 254
 - architecture, 265
 - business domain needs, 257
 - business scenarios, 256
 - challenges and best practices, 261
 - definition, 254
 - objectives and trends, 254
 - roles, 260
- digital commerce, 14
- digital content management, 28

- digital opportunities, 4
- distributed denial of service, 330
- DITA, *see* Darwin Information Typing Architecture
- document management, 263
 - capabilities, 263
 - case study, 267
 - evolution and roadmap, 266
- document type definition, 240
- DoS, *see* denial of service
- DTD, *see* document type definition
- Dublin core, 210

- ECM, *see* enterprise content management
- ECM capabilities, 17
 - content management, 16
 - document management, 16
 - enterprise services, 18
 - workflow, 18–19
- effective content design, 68
- enterprise content management, 14–15, 17
 - brief concepts, 15
 - capabilities, *see* ECM capabilities
 - comparison with WCM, 19
- enterprise digital ecosystem, 5
 - capabilities, 6
 - challenges, 5
 - trends, 14
- enterprise digital strategy, 21
- enterprise search, 378
 - best practices, 382
 - business drivers, 378
 - capabilities, 389
 - challenges, 381
 - comparison with web search, 384
 - evolution, 388
 - overview, 383
 - technology drivers, 380
 - trends, 386
 - value proposition, 389
- enterprise search features, 392
 - advanced search features, 393
 - basic search features, 392

- faceted navigation, 383
- federated search, 398
 - architecture, 399
 - features, 399

- HTML, 239
- information architecture, 196
 - best practices, 198
 - defining, 196
 - elements, 196
 - goals, 196
- information conservation, 24
- information consolidation, 23
- information value optimization, 23
- integrated experience, 23

- JavaScript Object Notation, 244
- JCR, *see* Content Repository API for Java
- JSON, *see* JavaScript Object Notation
- JSR 170, 235
- JSR 283, 235

- keyword search, 392
- learning management systems, 247

- messaging strategy, 34
- metadata, 75, 78, 102, 108, 113, 120, 127–128, 158, 162, 203
 - advantages, 214
 - application level metadata, 205
 - case study, 212
 - content personalization, 214
 - hierarchy modeling, 205
 - standards, 210
 - system-level metadata, 205, 215
 - types, 205
 - utilities, 214
- metadata management system, 222
- migration framework, 290
- MMS, *see* metadata management system
- mobile search, 386
- modern CMS, 106
 - architecture patterns, 106
 - microservices architecture, 111
 - MVC architecture, 106
 - N-tier architecture, 108
 - service-oriented architecture, 109

- omni-channel publishing, 62
- open source CMS, 100

- penetration testing, 338
- people search, 388

- permission inheritance, 336
- permission model, 334
- persona analysis, 44, 50
- personalized search, 404
- presentation template, 155
- promotion process, 131

- RDF, *see* resource description framework
- Really Simple Syndication, 246
- relevancy rank, 400
 - adjustment, *see also* search relevancy, 403
- representational state transfer, 239, 246, 249
- resource description framework, 378
- responsive web design, 65–66
- REST, *see* representational state transfer
- RSS, *see* Really Simple Syndication
- RWD, *see* responsive web design

- SCORM, *see* Sharable Content Object Reference Model
- search collection, 395–396
- search crawler, 384
- search engine optimization, 417
 - anti-patterns, 421
 - best practices, 420
 - page-level SEO, 416
 - SEO strategy, 420
- search facet, 393
- secured search, 405
- security code reviews, 337
- security testing, 337
- semantic metadata, 63–64
- semantic search, 409
- SEO, *see* search engine optimization
- Sharable Content Object Reference Model, 244
- shard, 417
- Simple Knowledge Organization System, 211
- Simple Object Access Protocol, 229, 235, 247, 248
- single sign-on, 334
- SKOS, *see* Simple Knowledge Organization System

- SOA, *see* adjustment, *see* service-oriented architecture
- SOAP, *see* Simple Object Access Protocol
- social search, 412
- social tagging, 217
- SQL injection, 329
 - prevention, 330
- SSO, *see* single sign-on

- taxonomy, 202
 - advantages, 203
 - best practices, 204
 - business drivers, 204
- template, 35, 83, 85, 90, 92, 95, 129, 132, 138, 154, 168–169, 205, 208
 - authoring template, 157
 - case study, 170
 - CMS support, 178
 - design, 154
 - development, 156
 - guidelines and best practices, 176
 - presentation template, 159
- TMS, *see* translation management system
- touch point optimization, 21
- translation management system, 226

- user journey modeling, 50

- WebDAV, *see* Web Distributed Authoring and Versioning
- Web Distributed Authoring and Versioning, 320
- web editorial calendar, 67
- wildcard search, 394
- workflow design, 184
 - case study, 189
 - CMS support, 189
 - design, 187
 - pre-requisites, 186
 - workflow optimization, 188

- XHTML, 239
- XML Schema Definition, 240
- XSD, *see* XML Schema Definition
- XSS, *see* cross-site scripting