Microsoft

Module 8: Implementing Search and Indexing

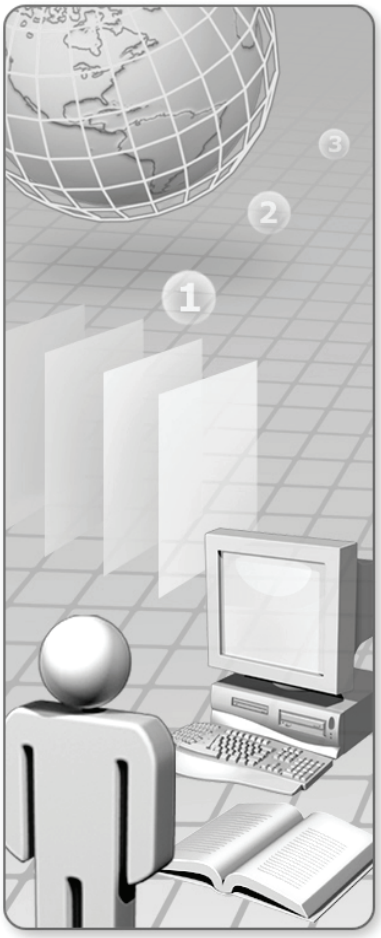


Table of Contents

Overview 8-1
Lesson 1: Office SharePoint Server 2007 Search and Indexing 8-2
Lesson 2: Implementing Search 8-7
Lab: Implementing Search and Indexing 8-19
Lab Shutdown 8-25
Review 8-26

Information in this document, including URL and other Internet Web site references, is subject to change without notice. Unless otherwise noted, the example companies, organizations, products, domain names, e-mail addresses, logos, people, places, and events depicted herein are fictitious, and no association with any real company, organization, product, domain name, e-mail address, logo, person, place, or event is intended or should be inferred. Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

The names of manufacturers, products, or URLs are provided for informational purposes only and Microsoft makes no representations and warranties, either expressed, implied, or statutory, regarding these manufacturers or the use of the products with any Microsoft technologies. The inclusion of a manufacturer or product does not imply endorsement of Microsoft of the manufacturer or product. Links are provided to third party sites. Such sites are not under the control of Microsoft and Microsoft is not responsible for the contents of any linked site or any link contained in a linked site, or any changes or updates to such sites. Microsoft is not responsible for Webcasting or any other form of transmission received from any linked site. Microsoft is providing these links to you only as a convenience, and the inclusion of any link does not imply endorsement of Microsoft of the site or the products contained therein.

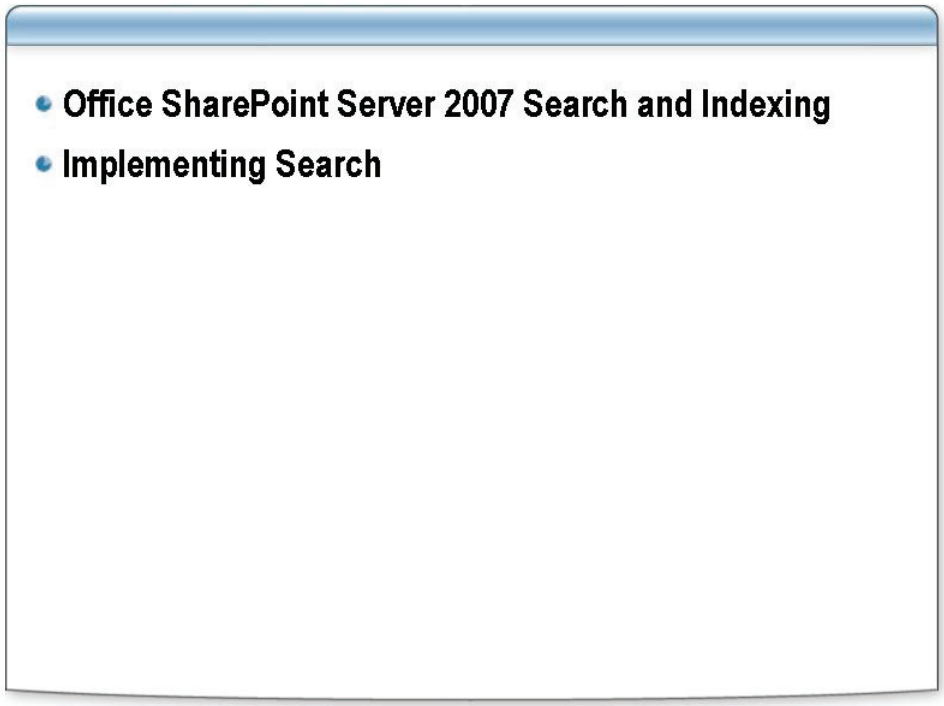
Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

© 2007 Microsoft Corporation. All rights reserved.

Microsoft, Active Directory, BizTalk, Excel, ForeFront, FrontPage, InfoPath, Internet Explorer, MSN, OneNote, Outlook, PowerPoint, SharePoint, SQL Server, Visual Basic, Visual C#, Visual Studio, Windows, Windows Server are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

The names of actual companies and products mentioned herein may be the trademarks of their respective owners.

Overview



Objectives

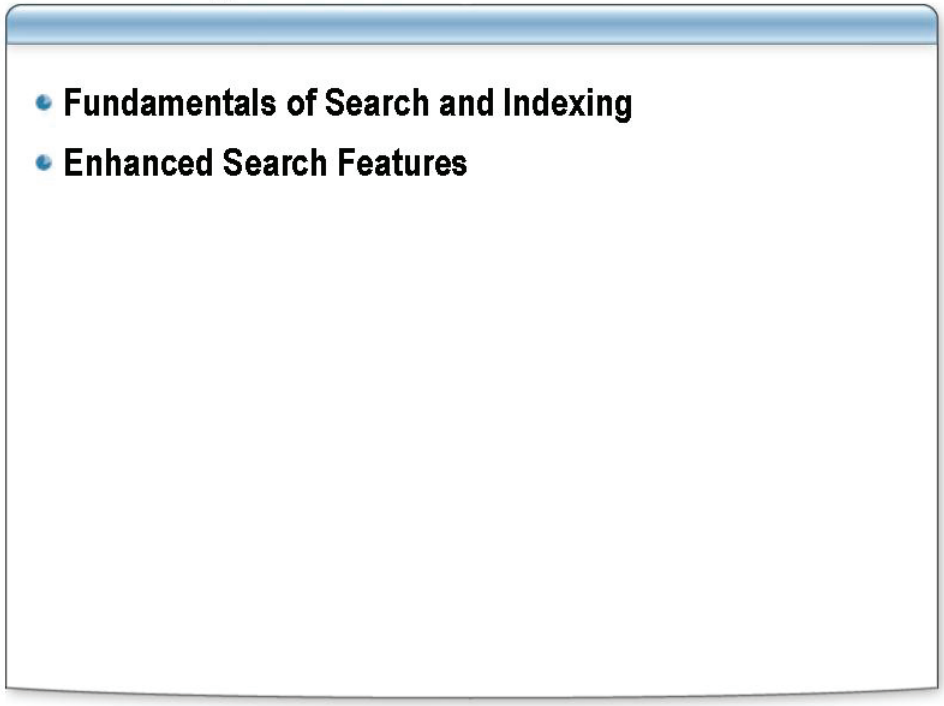
Organizations hold information in locations such as file shares, Web sites, business applications, and e-mails. Business users often require access to this data. The amount of information can grow everyday, particularly in the case of e-mails. Therefore, users must be able to search through vast amounts of information effectively. Microsoft® Office SharePoint® Server 2007 provides facilities for indexing and searching this data, which enables business users to work more efficiently. After completing this module, you will be able to:

- Explain the search and indexing features of Microsoft Office SharePoint Server 2007, and explain which features are provided by Windows® SharePoint Services and which are provided by Office SharePoint Server 2007.
- Describe the process of defining content to be indexed, indexing that content, and performing searches.

MCT USE ONLY. STUDENT USE PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

Lesson 1: Office SharePoint Server 2007 Search and Indexing



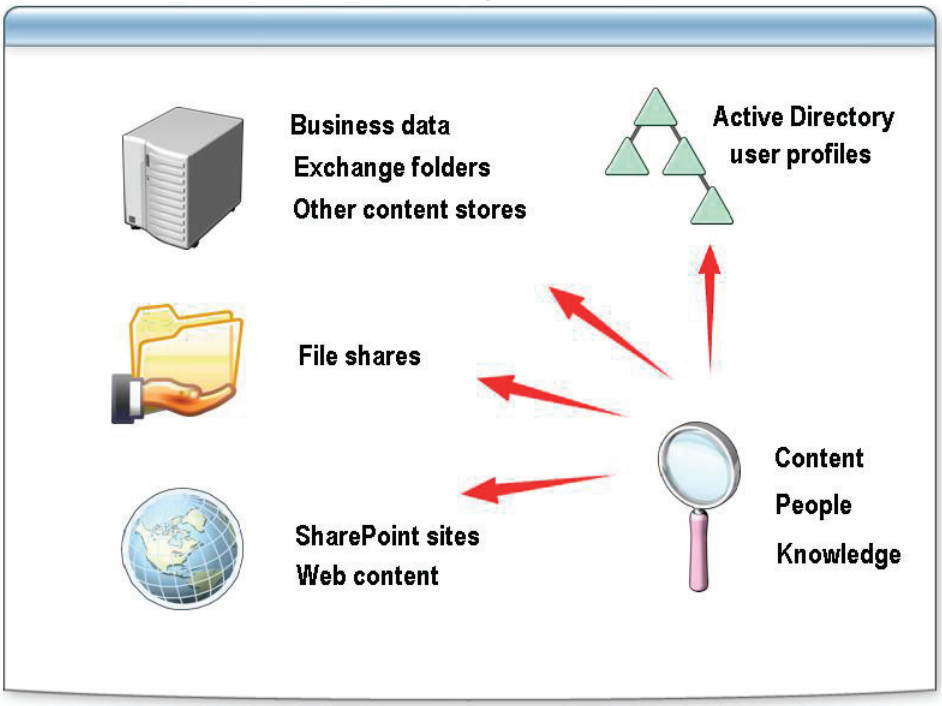
A search service must return relevant content to enable users to locate information distributed across their organization. This lesson explains the search and indexing features of Office SharePoint Server 2007. It also explains which features are provided by Windows SharePoint Services and which are provided by Office SharePoint Server 2007.

Objectives

After completing this lesson, you will be able to:

- Describe the enterprise search capabilities of Office SharePoint Server 2007.
- Describe the enhanced features of Office SharePoint Server 2007 Search.

Fundamentals of Search and Indexing



Enterprise Search Capabilities

You can configure Office SharePoint Server 2007 Search to crawl and index content stored in various locations in an organization. Locations include files shares and information stored in line-of-business applications.

Content

You can refine search queries in several ways to enable you to find the most relevant information. For example, you can use predefined content scopes to narrow the scope of your search or use advanced filters on content metadata, such as author and file type.

People

A key component of Microsoft Office SharePoint Server 2007 search is the ability to search for other people in the organization. You can search for people by skills or department. When you have found them, you can also identify their subordinates and managers.

Knowledge

Knowledge Network is an add-on to Office SharePoint Server 2007 that enables you to search for people by automating the discovery of the business relationships and subject matter expertise of everyone in the network.

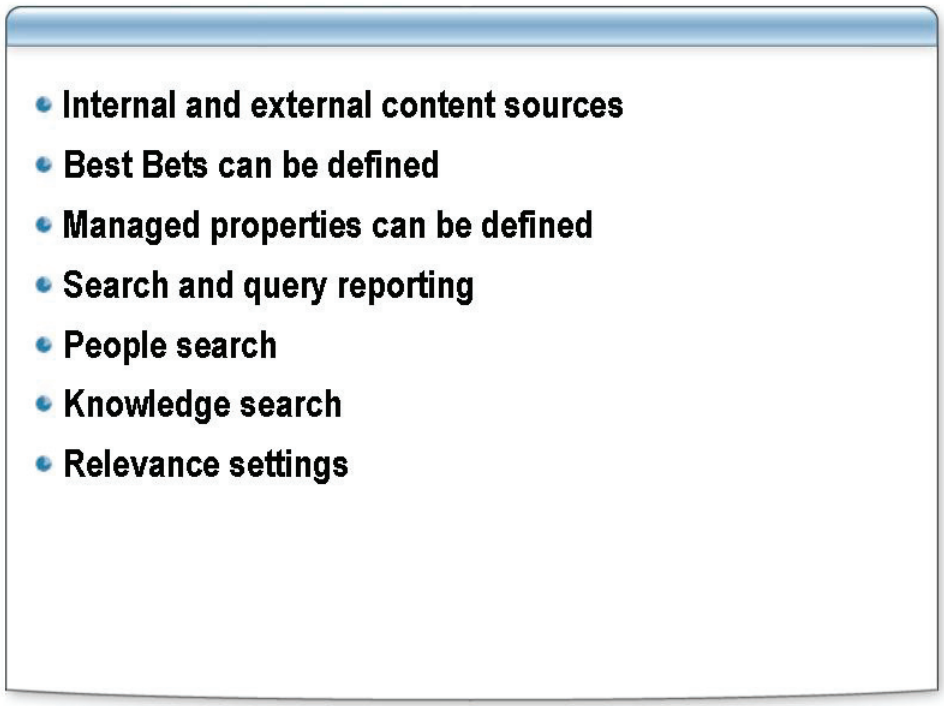
Key information is also stored in line-of-business applications. Office SharePoint Server 2007 Search can also search line-of-business application content exposed by the Business Data Catalog.

MCT USE ONLY. STUDENT USE PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

Note: Unlike previous versions of SharePoint Products and Technologies, Office SharePoint Server 2007 provides two search services: Windows SharePoint Services Help Search and Office SharePoint Server Search. You can use each of these services to crawl, index, and query content. Each service uses a separate index. If you do not want users to be able to search the Help system, you should not start this service.

Enhanced Search Features



Introduction

Windows SharePoint Services and Office SharePoint Server 2007 both use the same technology for search and indexing. Office SharePoint Server 2007 Search has several enhanced features to provide relevant search results. The enhanced capabilities of Microsoft Office SharePoint Server 2007 Search include the following options:

- Configuration of Office SharePoint Server 2007 Search to crawl internal and external content sources.
- Definition of Best Bets, which match popular keywords with the most relevant portal site pages, documents, or external Web sites.
- Definition of managed properties for advanced searching.
- Implementation of search and query reporting.
- The ability to search for people.
- The ability to search for knowledge.

Relevance

At the Shared Service Provider level, you can specify relevance settings for content locations to influence their position in search results. The relevance settings indicate how authoritative the content in a particular location is. There are four authoritative groups, as follows:

- Most authoritative pages.

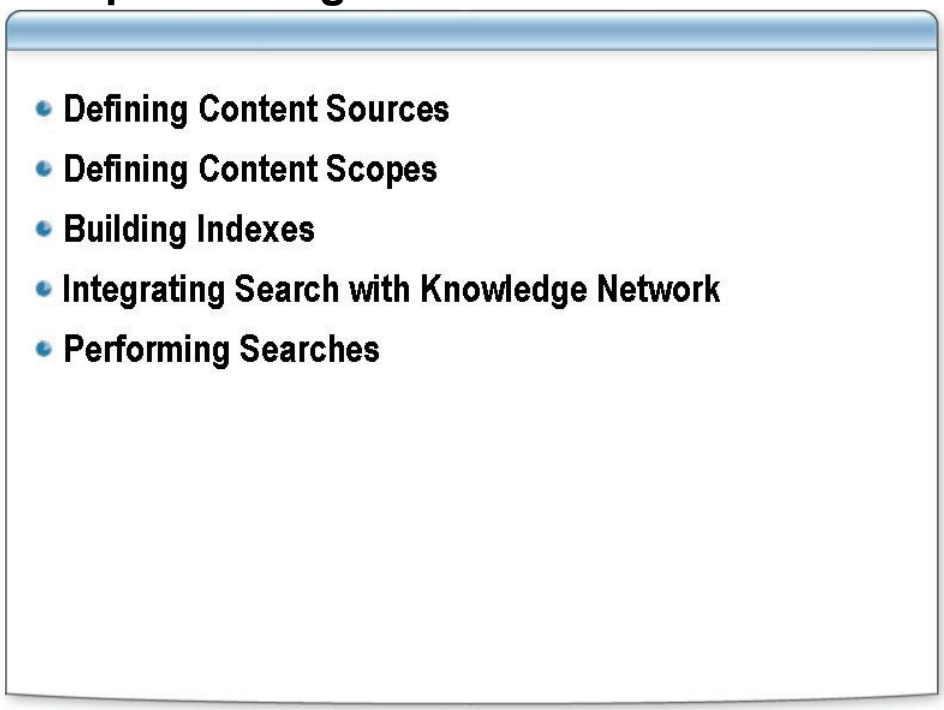
MCT USE ONLY. STUDENT USE PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

- Second-level authoritative pages.
- Third-level authoritative pages.
- Non-authoritative sites.

The relevance settings weight content in search results based on their authoritative level. However, the relevance settings do not supersede the keyword settings at the site collection level.

Lesson 2: Implementing Search



Office SharePoint Server 2007 Search enables administrators and information workers to easily configure and locate content stored across the organization. This lesson describes the process you use to index and locate content. You use this process to define the content you want to index, index that content, and perform searches.

Objectives

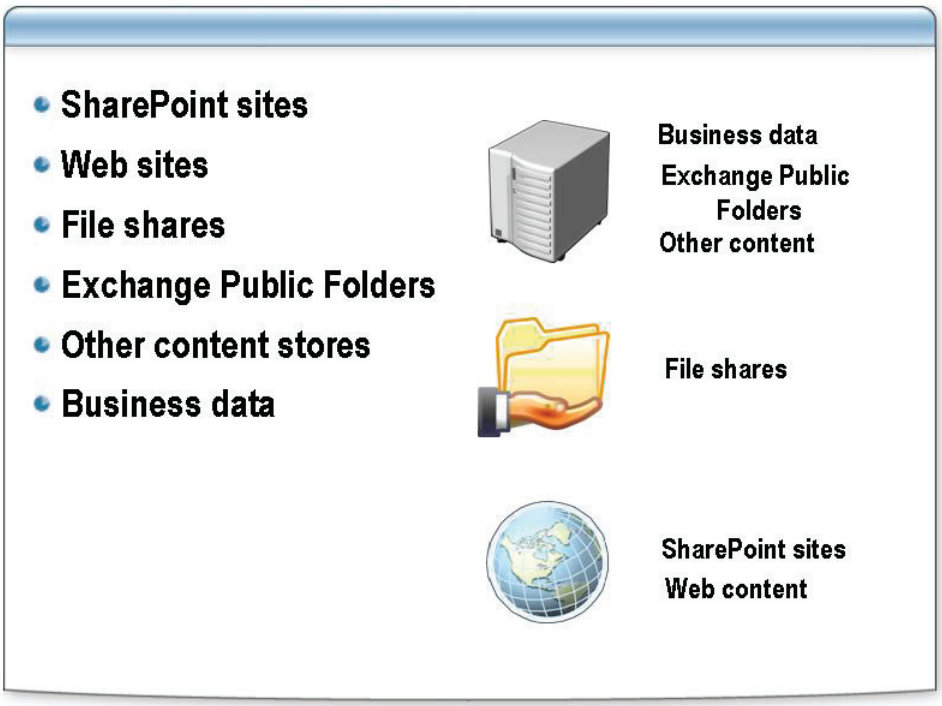
After completing this lesson, you will be able to:

- Explain how to identify and create content sources.
- Explain how to identify and assign content scopes.
- Describe how to build indexes.
- Describe the integration of Office SharePoint Server 2007 Search with Microsoft Knowledge Network.
- Describe how to configure and perform searches.

MCT USE ONLY. STUDENT USE PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

Defining Content Sources



Identifying Content Sources

You create content sources to identify the groups of related content or data that you want to crawl and index. Additionally, you must specify the crawl schedule for that content. Office SharePoint Server 2007 Search has one content source configured by default. This source includes all of the Web applications, and therefore all of the site collections, that use the Shared Service Provider.

You can create content sources for the following types of content:

- SharePoint sites, from the same server farm or different server farms.
- Web sites, internally or on the Internet.
- File shares.
- Exchange Public Folders.
- Other third-party content stores.
- Business data stored in line-of-business applications.

You should consider the following guidelines when you identify content sources:

- Group closely related content.
- Group content that is available at similar times.
- You can use a content source as the focus of a content scope rule.
- You cannot group business data content with other content in the same content source.

- Set crawler impact rules for external content sources that are not under your direct control.
- You can use a particular content source to crawl only one of the above content types.

Creating Content Sources

You define content sources at the Shared Service Provider level. To create a content source, perform the following steps:

1. Start **SharePoint 3.0 Central Administration**.
2. On the **Central Administration** page, click **Shared Services Provider**.
3. On the **Shared Services Provider** page, click **Search settings**.
4. On the **Configure Search Settings** page, click **Content sources and crawl schedules**.
5. On the **Manage Content Sources** page, click **New Content Source**.
6. Specify the **Content Source Type**, the **Start Addresses**, and the **Crawl Schedules** and then click **OK**.

MCT USE ONLY. STUDENT USE PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

Defining Content Scopes

How to Identify Scopes

Content scopes are named sets of rules that you use to filter search results. Content scope rules define the subsets of content that the search returns. The subsets of content should follow a common theme that users are likely to want to search through.

For example, you might define content scopes for each division in your organization. You can select your division content scope when you perform a search to return sales information from only your division.

You can define content scope rules for the following:

- Web addresses
- Properties (such as Author = Lisa Miller)
- Content sources
- All content

When content matches a rule it can be included, excluded, or used to satisfy a required match condition when combined with other rules in the content scope.

Assigning Content Scopes

You can create content scopes at the Shared Service Provider level or the site collection level. Content scopes that you create at the site collection level are only available for that site collection. Content scopes that you create at the Shared Service Provider level are

known as shared content scopes and can be used in all site collections. Two shared content scopes are available by default:

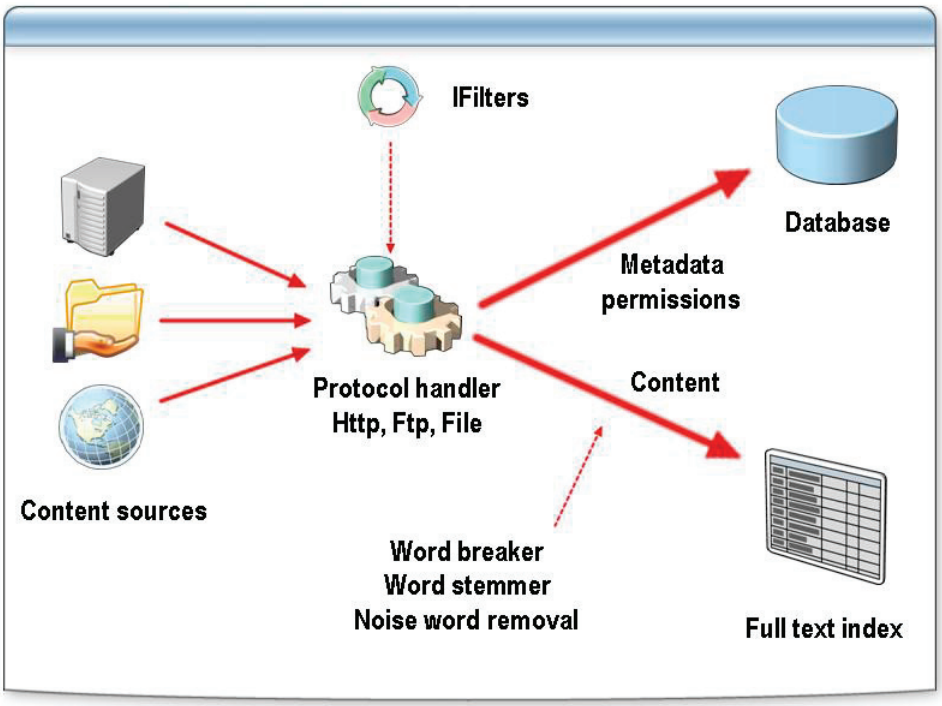
- All Sites
- People

You can make shared content scopes available to users at the site collection level. You can also copy and modify these scopes to create a unique content scope for a single site collection.

MCT USE ONLY. STUDENT USE PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

Building Indexes



Process

When Office SharePoint Server 2007 builds an index of a content source, the crawler uses an appropriate protocol handler to connect to the start address defined by the content source. The textual content is then processed. This stage uses any crawl and crawl impact rules that you have set. IFilters are the code that allows the Indexing Service to index different file formats and without which contents of a file cannot be indexed. The crawler will enlist suitable IFilters to extract text from any files it finds.

During the crawl, the crawler returns two streams, one for object metadata and permissions and another for object content. The properties and permissions are stored in a *property store table* and the content is sent to the content full-text index. The permissions information is used to filter the search results so that users see only content appropriate to their permission levels.

Before the content stream is put into the content index, word breakers and word stemmers are used to break the stream into words. Noise words are non-searchable words, such as "the" or "it". Because they are so common there is no value indexing them, so are also removed at this stage.

Note: Thesaurus files, word breaker files, and word stemmer files are installed in locale-specific subfolders of the C:\Program Files\Common Files\Microsoft Shared\web server extensions\12 folder by default.

The content index is then propagated from the index server to the query servers. The content indexes are automatically and continuously propagated. This ensures that the search results are as up-to-date as possible.

Crawler Settings

You can configure crawler settings at the server farm level, the server level, and the Shared Service Provider level.

Server farm-level settings include the following:

- The contact e-mail address that the crawler uses when crawling content.
- Proxy server details.
- Connection timeouts.

Crawler impact rules are also configured at the server farm level. These rules give you control over the number of simultaneous requests made for a particular site being crawled.

SSP-level settings include the following:

- The default content access account.
- Crawl rules for sites. You can configure the crawler to follow links from a site, but not to crawl the content on that site. You can also specify whether the crawler can follow links with query strings.
- The types of files that will be crawled, specified by file type extension.

Note: To search the contents of a file, you must install the appropriate IFilter.

Index Reports

At the Shared Service Provider level, reports are available for analyzing search queries and search results. These reports are critical for the continuous assessment of enterprise searches and will help you to identify areas for improvement.

The search queries reports are as follows:

- Queries over previous 30 days.
- Queries over previous 12 months.
- Top query origin site collections over previous 30 days.
- Queries per scope over previous 30 days.
- Top queries over previous 30 days.

The search results reports are as follows:

- Search results destination pages.
- Queries with zero results.
- Most clicked Best Bets.

MCT USE ONLY. STUDENT USE PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

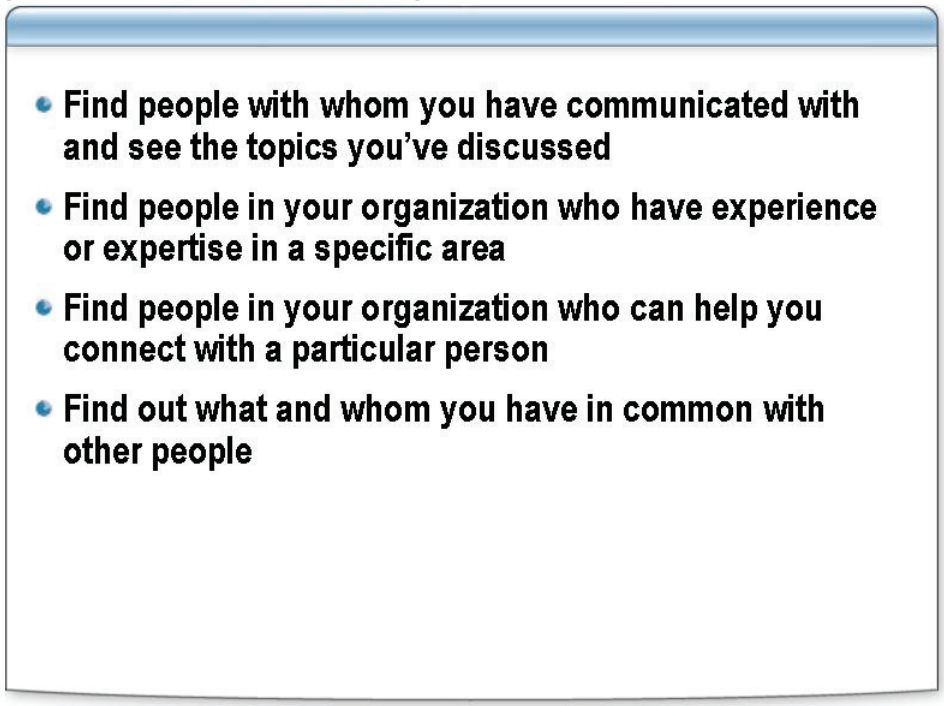
- Queries with zero Best Bets.
- Queries with low click through.

Performance Implications

Crawling content and building indexes are processor-intensive activities. You can use server-level settings to control the impact that indexing and crawling has on performance. For a particular server you can use the following settings:

- Enable or disable indexing.
- Limit the indexing activity for the server by limiting the number of threads that it can use.
- Designate a specific server for crawling content.

Integrating Search with Knowledge Network



Introduction

You can enhance Office SharePoint Server 2007 Search by opting to integrate it with Knowledge Network.

Knowledge Network consists of a client component and a server component. The Knowledge Network client is deployed on users' desktops. It crawls contacts and keywords from Microsoft Outlook®, Microsoft MSN® Messenger, Microsoft Windows Messenger, and Office Communicator. Keywords are determined from the subject and body of e-mails only, not attachments. The Knowledge Network client analyzes the data and recommends the best possible Knowledge Network profile for you, which you can then customize. This profile is synchronized with your My Site profile, when it is published to the Knowledge Network server component.

The Knowledge Network server component enhances Office SharePoint Server 2007 Search in the following ways:

- My Site is periodically updated with your keywords and contacts.
- You can perform more refined searches on expertise.
- You can search for external contacts (an External People tab is added to Search Center).
- You can find out who knows whom internally and externally (a Find People Who Know This Person link is made available on people in Search Center and My Site).

MCT USE ONLY. STUDENT USE PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

- You can use permissions to hide results from user searches, but let search show a link to request further information.

Knowledge Network enables you to perform the following tasks:

- Find people with whom you have communicated.
- Find people in your organization who have experience or expertise in a specific area.
- Find people in your organization who can help you connect with a particular person.
- Find out what and whom you have in common with other people.

Performing Searches

Searchable Entities

You can search for almost any content. The most common entities searched against are documents, Web page content, list item content, and people within an organization. You can also perform searches can against Business Data Catalog information.

Keywords

Keywords are specific words or phrases that you define at the site collection level to provide more targeted search results. A keyword can be associated with synonyms for a specific word or phrase, a definition of the keyword, and Best Bet links to content that is closely related to the keyword.

When you perform a search on the keyword or any of the synonyms defined for that keyword, the definition and any Best Bets will be displayed in the search results.

Note: Content scopes do not filter out any keyword results returned.

Basic Searches

To perform a basic search, use the search box on any page or in Search Center. To narrow a basic search, select a content scope in the Content Scope list if available.

Advanced Searches

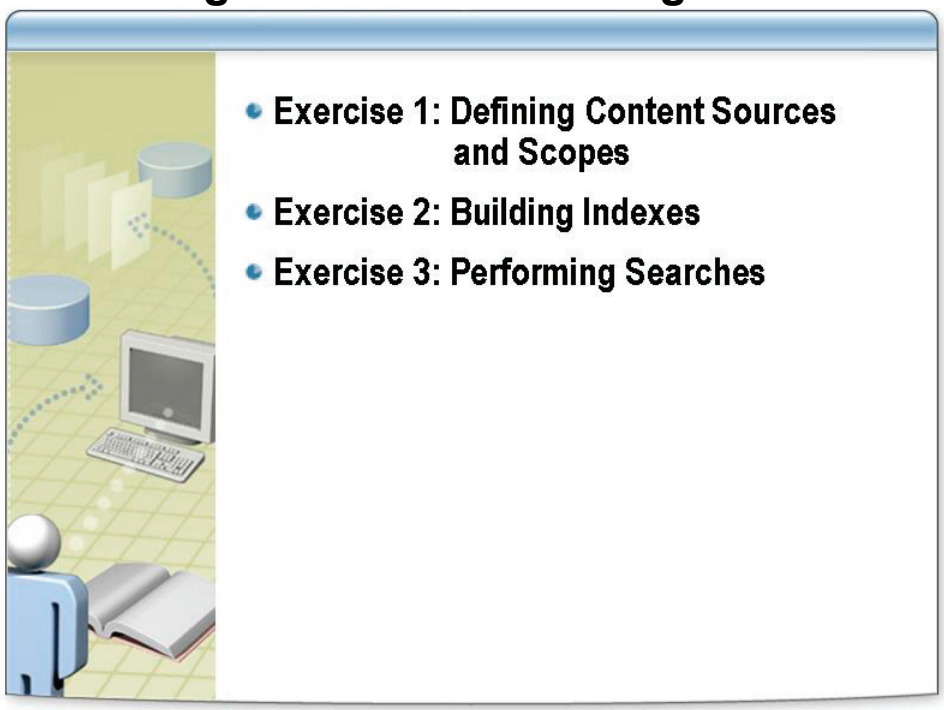
Advanced searches allow you to create a more targeted search. The Advanced Search page allows you to specify various refinements, including the following:

MCT USE ONLY. STUDENT USE PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

- Match an exact phrase or any, all, or none of the words entered.
- Narrow the search by properties and metadata.
- Narrow the search by document type.
- Filter the search by using content metadata.

Lab: Implementing Search and Indexing



After completing this lab, you will be able to:

- Define content sources and scopes.
- Build indexes.
- Perform searches.

Estimated time to complete this lab: 60 minutes

Lab Setup

For this lab, you will use the available virtual machine environment. Before you begin the lab, you must:

- Start the **5061A_LON-DC-01-08** virtual machine.
- Log on to the virtual machine with the user name **Administrator** and the password **Pa$$w0rd**.

Lab Scenario

The enterprise architect has determined the organizational goals, overall vision, and scope for Office SharePoint Server 2007 Search and Indexing in an enterprise content management solution. Your role is to implement a prototype query server to test some of the organizational search requirements. You will configure search and indexing on file shares and perform basic and advanced searches, and a people search.

MCT USE ONLY. STUDENT USE PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

Exercise 1: Defining Content Sources and Scopes

In this exercise, you will create content sources for file shares. You will also create a content scope and define rules to specify what content will be included in the scope.

The principal tasks for this exercise are as follows:

- Create content sources.
- Create a content scope.
- Define the rules for the content scope.

▶ Create content sources for the Technical Documentation file shares

Tasks	Detailed Steps
1. Create a content source for the Technical Documents file share.	• On the **Start** menu, point to **All Programs**, point to **Microsoft Office Server**, and then click **SharePoint 3.0 Central Administration**. • In the navigation bar on the left side of the page, under **Shared Services Administration**, click **SharedServices1**. After a short period of time, the **Home** page for the shared service provider appears. • On the **Home** page, click **Search settings**. • On the **Configure Search Settings** page, click **Content sources and crawl schedules**. • On the **Manage Content Sources** page, click **New Content Source**. • On the **Add Content Source** page, in the **Name** box, type **Technical Documents File Share** • In the **Content Source Type** section, select **File Shares**. • In the **Start Addresses** section, type **\\litwareportal\Technical Documents** • In the **Crawl Schedules** section, under the **Full Crawl** list, click **Create schedule**. • In the **Type** section, click **Weekly**. • Select **Sunday**, and then click **OK**. • Under the **Incremental Crawl** list, click **Create schedule**. • Select **Daily**, and then click **OK**. • At the bottom of the **Add Content Source** page, click **OK**.
2. Create a content source for the SharePoint Technical Documents file share.	• On the **Manage Content Sources** page, click **New Content Source**. • On the **Add Content Source** page, in the **Name** box, type **SharePoint Technical Documents File Share** • In the **Content Source Type** section, select **File Shares**. • In the **Start Addresses** section, type **\\litwareportal\SharePoint Technical Documents** • In the **Crawl Schedules** section, in the **Full Crawl** list, click **At 12:00 AM every Sun of every week**. • In the **Incremental Crawl** list, click **At 12:00 AM every day**, and then click **OK**. • If the **AutoComplete** dialog box appears, click **No**.

Create the Tech Docs content scope

Tasks	Detailed Steps
1. Create the Tech Docs content scope.	• On the **Manage Content Sources** page, in the breadcrumb control, click **Search Settings**. • In the **Scopes** section, click **View scopes**. • On the **View Scopes** page, click **New Scope**. • On the **Create Scope** page, in the **Title** box, type **Tech Docs** and then click **OK**.

Define the rules for the Tech Docs content scope

Tasks	Detailed Steps
1. Add a rule to include content from the Technical Documents File Share content source.	• For the **Tech Docs** item, click **Add rules**. • On the **Add Scope Rule** page, in the **Scope Rule Type** section, select **Content Source**. • In the **Content Source** list, click **Technical Documents File Share**. • In the **Behavior** section, ensure that **Include** is selected, and then click **OK**. • Leave Internet Explorer running. You will use it in the next exercise.

STUDENT USE PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

Exercise 2: Building Indexes

In this exercise, you will configure crawl rules for a Web site and specify the default content access account. You will also perform a full crawl for each of the content sources to build an index.

The principal tasks for this exercise are as follows:

- Build the index.
- Review the crawl log.
- Update the search scopes.

▶ Build the index

Tasks	Detailed Steps
1. Start a full crawl of each content source.	• In the breadcrumb control, click **Search Settings**. • On the **Configure Search Settings** page, click **Content sources and crawl schedules**. • Point to **Local Office SharePoint Server Sites**, and in the list that appears, click **Start Full Crawl**. • Point to **Technical Documents File Share**, and in the list that appears, click **Start Full Crawl**. • Point to **SharePoint Technical Documents File Share**, and in the list that appears, click **Start Full Crawl**. • Next to the **New Content Source** link, click **Refresh** until the **Status** for each content source reads **Idle**. This might take up to 10 minutes.

▶ Review the crawl log

Tasks	Detailed Steps
1. View the contents of the crawl log.	• In the breadcrumb control, click **Search Settings**. • Click **Crawl logs**. • Click **litwareportal**. • Review the content that was crawled by clicking the right pointing arrow next to **1 – 50** until you have reviewed all 261 items.

▶ Update the Tech Docs scope

Tasks	Detailed Steps
1. Update scopes.	• In the breadcrumb control, click **Search Settings**. • In the **Scopes** section, click **Start update now**. • After a minute, press F5 to refresh the page. Ensure that the **Scopes needing update** item reads **0**.
2. Add the scope to the LitwarePortal site.	• In the **Address** bar, type **http://litwareportal/** and then press ENTER. • On the **Site Actions** menu, point to **Site Settings**, and then click

Tasks	Detailed Steps
	Modify All Site Settings. • In the **Site Collection Administration** section, click **Search scopes**. • Click **Search Dropdown**. • Select **Tech Docs**, and then click **OK**. • Click the **Home** tab. • Leave Internet Explorer running. You will use it in the next exercise.

STUDENT USE ONLY. REPRODUCTION PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

Exercise 3: Performing Searches

In this exercise, you will perform several searches. These searches will show you how to use content scopes and Search Center to refine your searches. You will also search for people within the organization and see their relationships with other people in the organization.

The principal tasks for this exercise are as follows:

- Perform basic searches.
- Perform an advanced search.
- Perform a people search.

▶ Perform basic searches

Tasks	Supporting information
1. Perform a basic search from the LitwarePortal home page.	• On the **Home** page, in the **Search** box, type **planning search** and press ENTER. • Review the results and the number of results returned.
2. Perform a basic search from the Document Center page.	• Click the **Document Center** tab. • On the **Document Center** page, in the **Search** box, type **planning search** and press ENTER. • Review the results and the number of results returned. The results are the same as those returned by the previous search.
3. Perform a basic search within the Tech Docs content scope from the Document Center page.	• Click the **Document Center** tab. • On the **Document Center** page, select **Tech Docs** in the list next to the **Search** box. • In the **Search** box, type **planning search** and press ENTER. • Review the results and the number of results returned. The results are for documents that are defined by the rules for the Tech Docs content scope. Fewer results are returned than from the previous searches.

▶ Perform an advanced search

Tasks	Supporting information
1. Start with a basic search.	• On the **Search** page, in the **Search** box, type **document libraries** and press ENTER. • Review the results and the number of results returned.
2. Narrow the results by performing an advanced search.	• Click the **Advanced Search** link next to the **Search** box. • On the **Advanced Search** page, in the **Narrow the search** section, in the **Result type** list, click **Word Documents**. • In the **Add property restrictions** section, in the **Where the Property** list, click **Name**. • In the box, type **moss** • Click **Search**.

Tasks	Supporting information
	• Review the results. Fewer results are returned than from the previous search.

Perform a people search

Tasks	Supporting information
1. Perform a people search.	• On the **Search** page, just above the **Search** box, click the **People** tab. • In the **Search** box, type **Administrator** and press ENTER. • Click the first result. You are redirected to your **My Site**. • In the **Colleagues** section, you will see an entry for **Student**. Click the **Student** link. You can now see the Student user's colleagues. • Close Internet Explorer.

Lab Shutdown

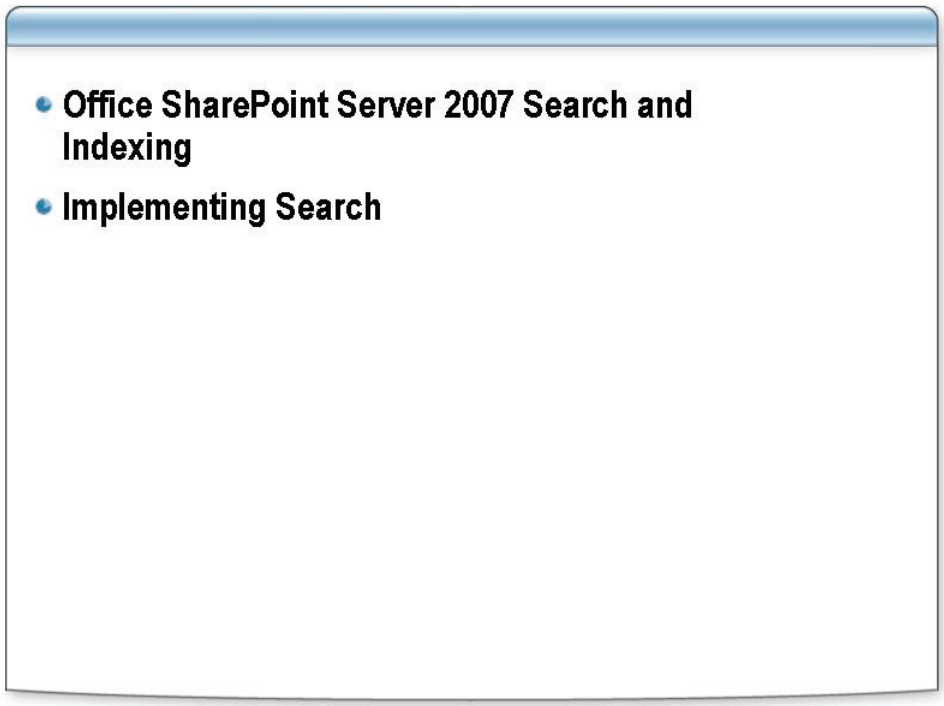
After you complete the lab, you must shut down the **5061A_LON-DC-01-08** virtual machine and discard any changes.

Important: If the **Close** dialog box appears, ensure that **Turn off and delete changes** is selected and then click **OK**.

T. STUDENT USE PROHIBITED

MCT USE ONLY. STUDENT USE PROHIBITED

Review



Review

You have now completed this module. You should now be able to:

- Explain the search and indexing features of Microsoft Office SharePoint Server 2007, and explain which features are provided by Windows SharePoint Services and which are provided by Microsoft Office SharePoint Server 2007.
- Describe the process of defining content to be indexed, indexing that content, and performing searches.