

Oracle® Fusion Middleware
User's Guide for Site Studio Publisher
11g Release 1 (11.1.1)
E10618-01

May 2010

Oracle Fusion Middleware User's Guide for Site Studio Publisher, 11g Release 1 (11.1.1)

E10618-01

Copyright © 2010, Oracle and/or its affiliates. All rights reserved.

Primary Author: Sean Cearley

Contributor: Brian Newton

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this software or related documentation is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, the following notice is applicable:

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation shall be subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License (December 2007). Oracle USA, Inc., 500 Oracle Parkway, Redwood City, CA 94065.

This software is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications which may create a risk of personal injury. If you use this software in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure the safe use of this software. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software in dangerous applications.

Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

This software and documentation may provide access to or information on content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services.

Contents

Audience.....	v
Documentation Accessibility	v
Related Documents	vi
Conventions	vi

1 Introduction

1.1 Understanding Site Studio Publisher.....	1-1
1.2 Installing Site Studio Publisher.....	1-2
1.3 Starting Site Studio Publisher	1-2
1.4 Guidelines for Publishing	1-2
1.5 Guidelines for Customizing Site Studio Pages.....	1-3
1.6 System Requirements	1-4

2 Publisher Tasks

2.1 About Publisher Tasks	2-1
2.2 Creating and Editing Tasks	2-1
2.3 Specifying Task Parameters.....	2-1
2.3.1 Settings	2-2
2.3.2 Dates	2-5
2.3.3 Days	2-6
2.3.4 Run Times	2-6
2.3.5 Options	2-6
2.3.6 Include List	2-7
2.3.7 Exclude List	2-7
2.3.8 Filter Sets.....	2-7
2.3.9 Triggers	2-8
2.4 Configuring Publisher.....	2-10

A User Interface

A.1 Site Studio Publisher Tasks	A-1
A.2 Add New Task/Edit Task	A-2
A.2.1 Publisher Settings	A-2
A.2.2 Dates	A-6
A.2.3 Days	A-6
A.2.4 Run Times	A-7

A.2.5	Options	A-7
A.2.6	Include List	A-8
A.2.7	Exclude List	A-8
A.2.8	FilterSets	A-9
A.2.9	Triggers	A-10
A.3	Task Info	A-12
A.4	Logs	A-13
A.5	View Log	A-13

Index

Preface

The *User's Guide for Site Studio Publisher* contains information to assist administrators responsible for the publishing of Web sites managed by Site Studio.

Audience

This document is intended for those people identified in the organization who are responsible for publishing an organization's Web site managed by Site Studio.

Documentation Accessibility

Our goal is to make Oracle products, services, and supporting documentation accessible to all users, including users that are disabled. To that end, our documentation includes features that make information available to users of assistive technology. This documentation is available in HTML format, and contains markup to facilitate access by the disabled community. Accessibility standards will continue to evolve over time, and Oracle is actively engaged with other market-leading technology vendors to address technical obstacles so that our documentation can be accessible to all of our customers. For more information, visit the Oracle Accessibility Program Web site at <http://www.oracle.com/accessibility/>.

Accessibility of Code Examples in Documentation

Screen readers may not always correctly read the code examples in this document. The conventions for writing code require that closing braces should appear on an otherwise empty line; however, some screen readers may not always read a line of text that consists solely of a bracket or brace.

Accessibility of Links to External Web Sites in Documentation

This documentation may contain links to Web sites of other companies or organizations that Oracle does not own or control. Oracle neither evaluates nor makes any representations regarding the accessibility of these Web sites.

Access to Oracle Support

Oracle customers have access to electronic support through My Oracle Support. For information, visit <http://www.oracle.com/support/contact.html> or visit <http://www.oracle.com/accessibility/support.html> if you are hearing impaired.

Related Documents

For more information, see the following documents in the Oracle Site Studio documentation set:

- *User's Guide for Site Studio Contributor*
- *User's Guide for Site Studio Designer*
- *Administrator and Manager's Guide for Site Studio*
- *Technical Reference Guide for Site Studio*

Conventions

The following text conventions are used in this document:

Convention	Meaning
boldface	Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.
<i>italic</i>	Italic type indicates book titles, emphasis, or placeholder variables for which you supply particular values.
monospace	Monospace type indicates commands within a paragraph, URLs, code in examples, text that appears on the screen, or text that you enter.

Introduction

This section covers the following topics:

- ["Understanding Site Studio Publisher"](#) on page 1-1
- ["Installing Site Studio Publisher"](#) on page 1-2
- ["Starting Site Studio Publisher"](#) on page 1-2
- ["Guidelines for Publishing"](#) on page 1-2
- ["Guidelines for Customizing Site Studio Pages"](#) on page 1-3
- ["System Requirements"](#) on page 1-4

1.1 Understanding Site Studio Publisher

This guide describes how to use Site Studio Publisher to publish a Site Studio Web site. With Site Studio, you store the Web sites you build in Oracle Content Server. At some point, you may find that you would like to create a static copy of a Web site and transfer it from an Oracle Content Server environment to a pure web server environment that is not running an Oracle Content Server instance. This process is referred to as *publishing*.

Site Studio Publisher creates a static snapshot of a dynamic site by traversing all the links in a Web site (visiting all of the linked pages) and downloading a copy of each page and all of the resources (images, flash movies, and so on) on each page. Your entire Web site, including the content of queries, layout pages, fragments, contributor data files, and native documents, are then copied and published to the new server.

The most efficient use of a dynamic site would have immediate updates and automatically update only the changed areas of the Web site. Since Site Studio Web sites are dynamic and created using a large number of variables, it is very difficult to track and publish the affected areas of the Web site in a reliable and consistent manner. Site Studio Publisher can selectively process, or not process, pages on a Web site based on different control methods. When each asset does not need to be compared each time, then it is much easier to keep the static site up to date, and the site owner can easily control the process.

Site Studio Publisher operates from an initial list of available links, and then compares them to a previous version of links used by Site Studio Publisher. The use of controls and filters can control the list, and also limit Site Studio Publisher to using only certain parts of the list. The first time Site Studio Publisher examines a Web site, all links are examined and added to the list.

When Site Studio Publisher examines a link and determines that it has changed, the content is analyzed and the new set of links for that page is determined. If the

examination is skipped, either because no changes were found or because the controls and filters require that a specific link is skipped, then the previous list of links is used for comparison.

1.2 Installing Site Studio Publisher

Site Studio Publisher installation is simple. On Oracle Content Server 11g, the Site Studio Publisher component is pre-installed, and is enabled through the Content Server UI.

To enable Site Studio Publisher, perform these tasks:

1. Log in to the Oracle Content Admin Server page with administrator credentials.
2. Go to the **Component Manager**.
3. Click the checkbox next to **Site Studio Publisher** under the **Web Content Management** section.
4. Click **Update** at the bottom of the page.
A dialog will appear confirming your selection.
5. The server must be restarted to complete enabling Site Studio Publisher.

To install and enable the Site Studio Publisher component on a 10gR3 Oracle Content Server instance, see the Content Server 10gR3 documentation.

1.3 Starting Site Studio Publisher

Site Studio Publisher is run from the Site Studio Administration section of the Oracle Content Server user interface.

To get to the Site Studio Publisher interface, perform these tasks:

1. Log in to the Oracle Content Server with administrator credentials.
2. Open the **Administration** tray or page, depending on your layout.
3. Select **Site Studio Administration**.
4. Select **Site Studio Publisher**.

1.4 Guidelines for Publishing

To publish a Web site from an Oracle Content Server environment to a web server, all the content of your site is copied by Site Studio Publisher. However, the copy of the site must be both complete and self-contained. That is, the copy must contain all of the resources of the original site, and pages in the copy must refer only to resources within the copy itself. To create a self-contained copy, Site Studio Publisher rewrites any absolute (full) URLs within the original site to relative URLs to allow the static copy to be hosted on a different hosting instance.

Site Studio Publisher uses the following procedures to identify and evaluate links and resources contained in a particular page:

- When handling HTML, Site Studio Publisher looks for any attributes in tags that may contain links (for example, the 'HREF' attribute in an 'A' tag or the 'SRC' attribute in an 'IMG' tag), and downloads these images. If the link is absolute, the link in the copied site is updated to refer to the downloaded image (through a relative URL) instead of the original image.

- When handling JavaScript, Site Studio Publisher looks for specific JavaScript commands that are commonly used to load a new image or change what is being displayed on the page. When those specific commands are encountered, it examines the JavaScript and interprets each command. However, an unrecognized JavaScript command may cause Site Studio Publisher to not include a referenced page or fail to download an image.

Be careful when appending URL parameters to a page URL. By default, in Site Studio Publisher, extra page parameters that result in a variant of the dynamic page are not used when naming the static page. Site Studio pages should not be customized by appending URL parameters to a node reference, as any appended parameters are ignored (that is, any node parameters not defined using Site Studio are not used in naming the pages).

1.5 Guidelines for Customizing Site Studio Pages

When customizing Site Studio-generated JavaScript or writing your own components, be sure to use relative URLs for all images, movies, and so on, that are referenced in the script; avoid concatenating strings to build up paths to resources. This makes it easier for Site Studio Publisher to properly identify and evaluate the external references.

Pay special attention to those items included in a dynamic list or using JavaScript patterns, and also to naming conventions for published pages.

Using JavaScript in Links for Published Pages

You may want to specify a JavaScript pattern in a link to be used in your published Web pages. Certain patterns must be used for Site Studio Publisher to identify and evaluate JavaScript references and links in order for these references to be converted to relative references in the published version.

Site Studio Publisher looks for specific patterns in the embedded JavaScript of Web pages. When these patterns are encountered, Site Studio Publisher examines the JavaScript and interprets the command.

- Site Studio Publisher evaluates general JavaScript patterns that are commonly used to load a new image or change what is being displayed on the page.
- Site Studio Publisher also evaluates certain Site Studio-specific JavaScript patterns.

Note: An unrecognized JavaScript command may cause Site Studio Publisher to not include a referenced page or fail to download an image.

In all patterns listed below, double quotation marks (") and single quotation marks (') are treated as equivalent. The crawler attempts to match any valid URL within the quotes.

General JavaScript Patterns

JavaScript fragment	Link to be traversed	Replacement
"http://www.example.com/test.html"	http://www.example.com/test.html	none

JavaScript fragment	Link to be traversed	Replacement
<code>something.src = "foo.gif";</code>	<code>foo.gif</code>	none
<code>top.location.href = "other.html";</code>	<code>other.html</code>	none
<code>foo.open("test.html");</code>	<code>test.html</code>	none
<code>frame.replace("something.html");</code>	<code>something.html</code>	none

Site Studio-Specific JavaScript Patterns

JavaScript fragment	Link to be traversed	Replacement
<code>var g_httpCgiUrl = "...";</code>		See note 1
<code>var ssAssetsPath = "..."</code>		See note 1
<code>var ssUrlPrefix = "..."</code>		See note 1
<code>link(XXX)</code>	<code>?IdcService=SS_GET_PAGE&ssDocName=XX</code>	See note 2
<code>nodelink(YYY)</code>	<code>?IdcService=SS_GET_PAGE&nodeId=YYY</code>	See note 2

Notes:

1. The parameters of these JavaScript calls, which are references to files or directories, are converted to a relative reference in the published version.
2. A hierarchical path is constructed using the URL Page Name and URL Directory Name properties of the site and the nodes. If no properties are defined the node label with underscores is used.

1.6 System Requirements

The Site Studio Publisher application requires Java version 5 (JRE 1.5) or higher.

Publisher Tasks

Site Studio Publisher allows you to define multiple tasks used in crawling and publishing the sites you maintain. These tasks are used to define how much of the site is crawled and how often the results are published.

Additionally, some default value of Site Studio Publisher, such as queue size, can be modified in the configuration file.

This section covers the following topics:

- ["About Publisher Tasks"](#) on page 2-1
- ["Creating and Editing Tasks"](#) on page 2-1
- ["Configuring Publisher"](#) on page 2-10

2.1 About Publisher Tasks

Tasks in Publisher are individual commands used to crawl a certain URL and then publish the gathered content. Each task can be made to crawl across an entire Web site, or across just a few pages.

Each task crawls the Web site (or portion of the Web site) at the time marked in the task. Tasks can also be created to run only on demand.

2.2 Creating and Editing Tasks

Creating a task requires little more than a URL of the site to crawl. Tasks are easy to create and edit because of the simple interface.

2.3 Specifying Task Parameters

The different parameters for a task are relatively straightforward. You need to have the URL for the Web site, and username and password, and you must set the parameters for what days the task will crawl the site. Some of the parameters, such as triggers, are more complex.

The parameters are grouped into separate sections on the page.

This section deals with the following topics:

- ["Settings"](#) on page 2-2
- ["Dates"](#) on page 2-5
- ["Days"](#) on page 2-6
- ["Run Times"](#) on page 2-6

- ["Options"](#) on page 2-6
- ["Include List"](#) on page 2-7
- ["Exclude List"](#) on page 2-7
- ["Filter Sets"](#) on page 2-7
- ["Triggers"](#) on page 2-8

2.3.1 Settings

The Settings of a task include some of the most basic information of a Publisher task. This includes the location of the site to crawl, any necessary login information and what errors are critical.

The Settings section has advanced settings, which can easily be seen when you click the **Show Advanced Settings** link. The advanced settings expand and display in the same section.

Description

The name of the task is used to differentiate it from other tasks. It can also contain some information about how often the task runs, or what site it crawls, and so forth.

Manifest Url

This is the URL for Site Studio Publisher to crawl and then publish. The server, instance, and site must be filled in by the user.

Output Path

The local directory where the content is downloaded. Important state information is stored here as well. This is relative to the *SSPHome* location.

Username

The user ID, for password-protected sites.

Password

The password, for password-protected sites. The password is encrypted in the *.hda* file.

Authentication

The type of authentication used on the site, for password-protected sites.

- **LoginForm** is used for Oracle Content Server 11g servers.
- **BasicAuth** is used for Oracle Content Server 10g servers.
- **ExtranetLook** is used for Oracle Content Server 10g servers with the ExtranetLook component.
- **NTLM** is used for Oracle Content Server 10g servers configured to use NTLM.
- **CustomForm** is used to customize a different form-based login.

Publish Now

Publish Now is a feature that marks pages with an implied high priority. As a contributor edits a Web page, they may mark a page by selecting "Publish Now" from Contributor. Publisher obtains a list of the marked pages from the server when it requests the list at the beginning of its operation.

Note: The "Publish Now" feature must be enabled in Designer. See the *User's Guide for Site Studio Designer* for more information.

When a crawl is executed, if the *Publish Now* control is set to true, Publisher only compares and publishes the marked, changed pages. If the altered content includes new links, Publisher also publishes those links. This crawl is suitable to run frequently.

Force Download

When Publisher typically examines links, the current server state is compared with the record of what Publisher previously found. What Publisher does not check is if the asset previously found is still in place on the local file system. The Force Download control is used to override in case the output file system has missing or corrupt assets.

Force Analyze

The Force Analyze control is useful when you have implemented filters in your task specification. This control is only invoked during analysis, so use of this control forces the analysis pass on those files marked for examination on this crawl. If no controls are used to limit the examination operation, all files are re-analyzed with this setting.

This control is intended for use during development. The best use of the control is when *filterset* and *filter* controls within the *task* specification are used. If a filter is changed, then this usually affects some link rewriting behaviors. Normally, if the dynamic page has not changed, the filters are not re-applied. The result is that when a dynamic page has not changed, the links are not re-written using the changed filters.

Treat Home Page Errors As Critical

This control allows you to stop the task if errors are found while retrieving the home page. If the box is checked, then any error on the home page will abort the task.

Treat Manifest Errors as Critical

This control allows you to stop the task if there are errors in the URLs listed in the site manifest. If the box is checked, then any errors found retrieving the URLs in the site manifest will abort the task.

Use Cache Control

Cache Control is a control that allows the server to compare the Max Age (caching) section property controlling the *CacheControl:max-age* http header. Using this control, Publisher only selects and crawls those pages where the max-age value has not expired.

This allows various parts of the site to be labeled as "rarely changed" and still create, control, and execute positive checks at controlled intervals. This enables you to run weekly, monthly, quarterly, or even annual publishing controls to enable timely execution of the publication of the Web site.

Use Last Modified

The Last Modified control is used to compare the http header *Last-Modified* returned from the web server. The web server normally returns this header for all resources accessed with a weblayout URL. (This response header is not provided for a dynamic Site Studio page.)

If *useLastModified* is checked, the *Last-Modified* value is re-submitted in an *If-Modified-Since* request header the next time that resource is retrieved. That then

allows the web server to return a *304 - Not Modified* response if the resource is unchanged. If this is not checked, the resource is downloaded and compared with the previously retrieved content.

You would only choose to uncheck this control if you found that your web server was returning unreliable results for *Last-Modified*.

Use SSPETag

The SSPETag control is used to identify changes to files retrieved by the GET_FILE service. Use of this mechanism allows the content server to return a *304 - Not Modified* response when applicable and so avoid unnecessary downloads.

Without this mechanism, file content is retrieved and compared with the previous version.

Default Filename

The default filename is the filename used for URLs that do not have an explicitly-specified filename.

Page Extension

The extension that is added to page urls that do not otherwise specify an extension. For example, a typical reference to a document in the dynamic site might look like:

```
http://myServer/mySite/Section1/DocumentX
```

Where DocumentX is the *dDocName* of the target item in Content Server. For use on a static site, you must add an extension.

User Agent

This control enables you to specify a value for the *User-Agent* http request header used by Site Studio Publisher when crawling the site.

Friendly Url Parameters

This control enables you to select a comma-separated list of additional parameter names to honor. Friendly Urls already honor parameters ending in *NextRow* or *dcPageNum*, and use the parameter name and value to construct the filename for the crawled page.

If a URL Parameter affects the appearance of the page, then you need to capture a different copy of the page for each combination of parameter values.

Dynamic Url Parameters

This control enables you to select a comma-separated list of additional parameter names to honor. Dynamic Urls using the GET_PAGE service already honor *dID*, *dDocName*, *RevisionSelectionMethod*, and *Rendition*.

If a URL Parameter affects the appearance of the page, then you need to capture a different copy of the page for each combination of parameter values.

Additional Services

This control enables you to select additional services to be crawled. There is built-in support for those services that are expected to generate meaningful static content: SS_GET_PAGE, GET_FILE, and GET_DYNAMIC_CONVERSION. Any services entered will be added to those three as the crawled services. There is not a method to prevent the crawl from using one of the three default services.

For more information on the Site Studio Services, see the *Technical Reference Guide for Site Studio*.

Soft Error Threshold

This control determines the number of soft errors allowed. If this number is exceeded, publishing fails.

Hard Error Threshold

This control determines the number of hard errors allowed. If this number is exceeded, publishing fails.

Delete Threshold

This control determines the number of missing or deleted objects found while crawling. If this number is exceeded, publishing fails. Entering -1 in this field allows any number of missing objects.

Soft Error Codes

This control enables to you specify which server error codes are treated as soft errors. No wild cards are allowed; each code must be entered explicitly in a comma-separated list.

Hard Error Codes

This control enables to you specify which server error codes are treated as hard errors. No wild cards are allowed; each code must be entered explicitly in a comma-separated list.

Ignore Error Codes

This control enables to you specify individual error codes that are ignored. These codes will not affect the Site Studio Publisher crawl. No wild cards are allowed; each code must be entered explicitly in a comma-separated list.

2.3.2 Dates

The scheduling of a task starts with the Dates section. Here, a task is set as either a task that runs according to a schedule, or is run only on demand.

If you select that a task is run on demand - that is, by selecting **Manual** - then the Days (see "[Days](#)" on page 2-6) and Run Times (see "[Run Times](#)" on page 2-6) sections will be unavailable.

Manual / Range

This control enables you to set if the task will be run only when selected, or if it will run at regular, scheduled intervals within the selected date range.

If you select Manual, then the *Start Date* and *End Date* fields, and the Days (see "[Days](#)" on page 2-6) and Run Times (see "[Run Times](#)" on page 2-6) sections will be made unavailable.

Start Date / End Date

This control enables you to set the start date and end date of the scheduling period for this task. There is a calendar control next to each date field.

2.3.3 Days

The Days section controls the scheduling of the task by day. Here you can set if the task is run on certain days of the week (such as every Monday and Thursday) or on certain days of the month (such as the 10th and 20th of the month).

The options in this section will be available only if **Range** was selected in the Dates section (see "[Dates](#)" on page 2-5).

Days of Week

This control enables you to specify which days of the week the task will run, regardless of what day of the month that day is.

Days of Month

This control enables you to specify which days of the week the task will run. This list can be a range, or a comma-separated list of specific dates, or a combination. For instance, 1, 15 and 10-20 and 1, 7-10, 15-19, 23 are all valid entries.

You can also use the word `last` in this field to denote the last day of the month. It can be used in a list, such as 15, `last` to run on the 15th of the month as well as the last day of the month.

2.3.4 Run Times

The Run Times section is used to specify how often the task runs each day.

The options in this section will be available only if **Range** was selected in the Dates section (see "[Dates](#)" on page 2-5).

Multiple Times per Day / Once per Day

This control enables you to select how often the task runs each day, and during what times of the day.

If you select Multiple Times per Day, you should then enter how often the task will run using the *Hours* and *Minutes* fields. Additionally, you must enter the time of day that the task will run during.

If you select Once per Day, the task will run only at one time of day, which is entered in the box.

2.3.5 Options

The Options section is used to set the priority and log levels of the task. You can also set an email address to receive notification after the task runs.

Priority

This control enables you to select if the task will be run at a priority level. If unchecked, the task will run at the normal level.

Priority tasks are run in a different queue than normal tasks. Since most tasks are expected to be set to normal priority, the expectation is that few, if any, tasks will be run with a high priority, enabling those that are set at high priority to run sooner and end sooner. Additionally, priority tasks would check only part of a site, by using *Publish Now* or *Include List*.

Log Level

This control enables you to set how much information is entered in the log.

Each item in the drop-down list includes the logging levels above it. For example, selecting **INFO** includes logs of not only **INFO** items but also **WARN** and **ERROR**.

Email Notification

This control enables you to enter an email address that Publisher will send an email message to once the task has completed (or failed due to errors).

2.3.6 Include List

This control enables you to select individual site URLs which will be specifically included in the task. If a URL entered here is also listed in an exclude filter in the Filter set (see "[Filter Sets](#)" on page 2-7), the filter set takes precedence, and the URL will **not** be included in publication.

Each time you click **Add New Item**, an additional field for entering a URL will display.

2.3.7 Exclude List

This control enables you to select individual site URLs which will be specifically excluded from the task.

Each time you click **Add New Item**, an additional field for entering a URL will display.

2.3.8 Filter Sets

A filter element describes a transformation that is applied to a downloaded content file. A filterset contains filter elements, and each filter is applied in the order the filter elements appear in the filterset. Several types of filter can be specified. Specification of a filter element depends on the kind of filter used, as specified by the type attribute.

Filtersets are used in different contexts: when applying filters to a downloaded file, for instance, or when deciding whether a given URL should be downloaded.

Each attribute of a filterset is compared to the relevant characteristics of the source content. Each and every attribute must match for the filterset itself to be considered a match. The filtersets are evaluated in the order in which they appear in the job description file, and the first filterset match is applied.

Type

This filterset attribute specifies the content that the filterset will invoke. The types available are:

- **transform-content:** for URLs that match this filterset, the child filter elements are applied during download, transforming the content in the manner specified by these filters.
- **transform-link:** for URLs that match this filterset, the child filter elements are applied during download. In this case, the filter elements are only applied to links found in the current downloaded file, not to the entire content.
- **exclude:** for URLs that would normally be included, match this filterset, and the URL matches, content is not downloaded.

Path

This filterset attribute specifies a wildcard pattern to match the file path of the URL (the part following the URL's host name) currently downloaded. This attribute uses Java regular expression syntax for pattern matching.

Use the following reference for Java syntax:

<http://java.sun.com/j2se/1.4.2/docs/api/java/util/regex/Pattern.html>

Host Name

This filterset attribute is a value used to match the URL's host name.

Port

This filterset attribute is a value used to match the URL's port number.

Mime Type

This filterset attribute is a value used to match the URL's MIME-type. The URL values that starts with the entered filter value are considered a match; `text` matches both `text/html` and `text/xml`.

Each filterset is a collection of at least one filter. These filters act in the order they are listed within the filterset, just as the filtersets act in the order they are listed.

The filters have the following attributes:

Regex

This control is the regular expression used to identify text to replace in the context of the file or link.

Replacement

This is the replacement value for the identified *regex* string.

Path Regex

This control is the regular expression used to identify a path and file to replace in the context of the current file. The **Path Regex** and **Path Replacement** are the best method of finding and renaming output files.

Using path regex is only meaningful within an enclosing *transform-content* filterset. Additionally, it does not use the **Global** and **Ignore Case** checkboxes. It is always case-sensitive, and replaces only the first match.

Path Replacement

This is the replacement value for the identified *path regex* string.

Global

This attribute is used to control the behavior of the substitution. If **Global** is checked, every match for the regex expression is replaced. If unchecked, only the first occurrence is replaced.

Ignore Case

This attribute is used to control the behavior of the *regex* expression. If **Ignore Case** is selected, then the case of the expression is ignored when comparing the *regex* value.

2.3.9 Triggers

Triggers enable the replication engine to run commands after the download is complete. A trigger is defined by a trigger element.

Username/Password

The username and password used for authentication. You cannot specify different authentications for each trigger. All triggers in a task must have the same credentials and authentication method.

Authentication

Select to authenticate the entered username and password either with basic authentication or with a custom form.

Type

The type control determines the type of command functionality.

A **cmd** trigger allows a shell script or an operating system command to be run as a trigger. The command is passed to the operating system.

The **http-post** trigger sends a command to the specified URL using HTTP_POST functionality.

The **http-get** trigger sends a command to the specified URL using HTTP_GET functionality.

The **http-soap** trigger sends a command, encoded as a SOAP package, to the specified URL using HTTP_POST functionality.

Command

This trigger control holds a command-line if the type is set to **cmd**. For the other **http** trigger types, the control holds a URL.

This is the command-line argument to pass through the trigger.

Examples of a command can be:

```
ping MyServer
cmd /C ren %%d\*.xml *.html
```

The first example (in Windows) would simply ping the MyServer server. The second is an example of a Windows command-line trigger renames any file with the filename extension of `.xml` to use the filename extension `.html`.

To run this trigger in the directory where content files have been downloaded, you must use `%%d`, followed by `\` in a Windows environment or `/` in a UNIX environment, to represent the name of the download directory.

Command Data

This control is used to specify the path, relative to the task's output path, of the file to be uploaded.

This is valid only when the trigger type is **http-post** or **http-soap**.

For the **http-post**, is the application/x-www-form-urlencoded data to be submitted with the POST. Essentially, this is a list of name-value pairs separated by the `&` character. For example:

```
objectType=package&actionType=update&packageName=NAME
```

For **http-soap**, it is the location of a file containing the SOAP commands to be executed.

Response File

This control is used to specify a partial path to the response file relative to the **Output Path** for the task. The response file is used to capture all trigger responses.

SOAPAction

This control is used to set the value for the SOAPAction HTTP request header field.

This is valid only when the trigger type is set to **http-soap**.

Run the trigger even if the crawl failed

This control is used to determine if the trigger command should execute, even if there are other errors in the crawl.

Run the trigger if there were no changes

This control is used to specify if the trigger will run if there has been no change to the site.

Ignore trigger failure

This control is used to determine if subsequent triggers should be run, or if the task should simply stop. If checked, further triggers will run even if the trigger fails.

Log the response

This control is used to specify if the trigger response is written to the logfile.

2.4 Configuring Publisher

There are some changes you can make to the default values of Publisher. The *config.cfg* file can be edited to allow different default actions and paths to be used.

Attribute	Purpose	Defaults
SSPHome=D:/SSPHome	This option sets the "home" directory for Publisher's relative paths for output and auxiliary files.	data/SiteStudioPublisher
SSPLockFileTimeout=30	This option controls the time that the Publisher output directory is locked. This value defines how old a file must be before it will be considered "stale" and be overridden.	30 (seconds)
SSPOverrideStaleLockFile=true	This option enables or disables the SSPLockFileTimeout mechanism.	true (enabled)
SSPDeleteLockFile=true	This option determines if Publisher will delete the "lock" file at the end of a run.	true

Attribute	Purpose	Defaults
SSPTimerTickInterval=5000	This option determines, in milliseconds, the interval at which Publisher will examine the task schedule and queues to run eligible tasks and update status	5000 (milliseconds)
SSPViewLogPageLength=5000	This option lets you control the number of lines displayed per page when viewing a log file. By default, Publisher breaks a log file into 5000 line chunks for viewing in the "View Log" page (see "View Log" on page A-13).	5000 (lines)
SSPLogFileAgeLimit=7	This option determines how long log files are stored. Publisher will keep log files spanning up to 7 days for each task. Old logs for a specific task are discarded whenever that task is run.	7 (days)
SSPPriorityTaskThreads=1	This option determines the number of tasks marked "Priority" that run at the same time. The sum of the <i>Priority</i> and <i>Normal</i> tasks should not exceed 4.	1
SSPNormalTaskThreads=1	This option determines the number of tasks not marked "Priority" that run at the same time. The sum of the <i>Priority</i> and <i>Normal</i> tasks should not exceed 4.	1
SSPProxyHost	This option lists the address of an http proxy server, if used. If set, this will override any JVM values used.	No default value. (Example: proxy.example.com)
SSPProxyPort	This option lists the port used for the http proxy server, if used. If set, this will override any JVM values used.	80
SSPNonProxyHosts	This option lists any servers that should be connected to directly and not through the proxy listed in <i>SSPProxyHost</i> .	No default value. (Example: MyServer *.example.com)

The *config.cfg* file is accessed in the **General Configuration** section of the Admin Server, on the **Additional Configuration Variables** screen.

User Interface

The Oracle Content Server interface includes various pages that you can use to manage the publishing process of your Site Studio Web sites:

- ["Site Studio Publisher Tasks"](#) on page A-1
- ["Add New Task/Edit Task"](#) on page A-2
- ["Task Info"](#) on page A-12
- ["Logs"](#) on page A-13
- ["View Log"](#) on page A-13

A.1 Site Studio Publisher Tasks

This screen lists the publishing tasks that have been defined, and details about their operation.

Figure A-1 Main Site Studio Publisher Screen

Running							
Pause Publishing		Add New Task		Refresh			
Description	Priority	Status	Next Run	Last Run	Progress	Actions	
Weekly Tuesday Automated Publish	normal	waiting	Nov 17, 2009 7:00:00 AM				
runondemand	high	waiting		Nov 16, 2009 2:43:19 PM	Completed		

Element	Description
Pause Publishing	Click to stop the scheduling process from starting new tasks. This will pause scheduling new tasks, but will not pause tasks that are currently running. This button only appears when tasks are not paused.
Resume Publishing	Click to allow the scheduling process to start new tasks. This button only appears when tasks are paused.
Add New Task	Opens the Add New Task screen (see "Add New Task/Edit Task" on page A-2) to enter a task for publishing.
Refresh	Refreshes the list of tasks.
Description	The name of the task.

Element	Description
Priority	Displays if the task runs as a high priority or normal priority.
Status	Displays if the task is Waiting, Pending, Running, Finished, or Expired .
Next Run	Scheduled date and time that the task will run.
Last Run	Date and time of the completion (or failure) of the most recent run of the task. If this field is blank, the task has not been run since the last time the server was restarted.
Progress	Current progress status of the task.
Actions	Click to open a menu and select an action for the task. Edit: opens the Edit Task screen (see "Add New Task/Edit Task" on page A-2) to make changes to the task. Run: puts the task in the queue to run, and will continue to run according to the schedule set for the task. This is the only method to run manual tasks. Stop: available only while a task is running. Stops the task. Delete: deletes the task. View Info: opens a screen to view the information about previous times the task has run. This is disabled on tasks that have never run. View Logs: opens the screen to view the logs for the task.

A.2 Add New Task/Edit Task

The page used for adding and editing tasks to run with publisher has many parts to it. It is here that you can specify how often you publish a site, what site to publish, and even which sections of the site to include or exclude as well as other filters. Triggers can be set as well.

This section covers the following topics:

- ["Publisher Settings"](#) on page A-2
- ["Dates"](#) on page A-6
- ["Days"](#) on page A-6
- ["Run Times"](#) on page A-7
- ["Options"](#) on page A-7
- ["Include List"](#) on page A-8
- ["Exclude List"](#) on page A-8
- ["FilterSets"](#) on page A-9
- ["Triggers"](#) on page A-10

A.2.1 Publisher Settings

The Publisher Settings section is used to set the basic information of the task. The location of the site to publish, the location of the output, the username and password, and so forth.

The Settings section has advanced settings, which can easily be seen when you click the **Show Advanced Settings** link. The advanced settings expand and display in the same section.

Figure A-2 Publisher Settings Section (Basic)

Settings

Description

Manifest Url

Output Path

Username

Password

Authentication ▼

[Show Advanced Settings](#) Publish Now

Figure A-3 Publisher Settings Section (Expanded)

Settings

Description

Manifest Url

Output Path

Username

Password

Authentication ▼

[Hide Details](#) Publish Now

Force Download

Force Analyze

Treat home page errors as critical

Treat manifest errors as critical

Use Cache Control

Use LastModified

Use SSPETag

Default Filename

Page Extension

User Agent

Friendly Url Parameters

Dynamic Url Parameters

Additional Services

Soft Error Threshold

Hard Error Threshold

Delete Threshold

Soft Error Codes

Hard Error Codes

Ignore Error Codes

Element	Description
Description	Enter a name of the task.

Element	Description
Manifest Url	Enter the URL for Site Studio Publisher to crawl. This is a specially formatted url that gets information about the Web site to initiate the crawl.
Output Path	Enter the path to the local directory where content is downloaded. This is relative to the <i>SSPHome</i> location.
Username	Enter the user ID for password-protected sites
Password	Enter the encoded password for password-protected sites.
Authentication	Select the type of authentication for the site. <ul style="list-style-type: none"> ■ LoginForm is used for Oracle Content Server 11g servers. ■ BasicAuth is used for Oracle Content Server 10g servers. ■ ExtranetLook is used for Oracle Content Server 10g servers with the ExtranetLook component. ■ NTLM is used for Oracle Content Server 10g servers configured to use NTLM. ■ CustomForm is used to customize a different form-based login.
Publish Now	Select to compare and publish only the marked, changed pages. If the altered content includes new links, Site Studio Publisher also publishes those links.
Show Advanced Settings	Click to display the Advanced Settings.
Hide Details	Click to collapse the Advanced Settings.
Force Download	Select to force a download of the Web page whether there are found changes or not.
Force Analyze	Select to force the analysis pass of the files listed in the filter parameters. If no files are listed in the filter, then all pages are analyzed. This is used only with filters.
Treat home page errors as critical	Select to specify if any error retrieving the home page aborts the job.
Treat manifest errors as critical	Select to specify if any error retrieving any of the page urls listed in the site manifest aborts the job.
Use Cache Control	Select to compare the Max Age section property. Site Studio Publisher only selects and crawls those pages where the max-age value has not expired.
Use Last Modified	Select to use the value of the http header <i>Last-Modified</i> returned from the web server. The web server normally returns this header for all resources accessed with a weblayout URL. (This response header is not provided for a dynamic Site Studio page.) If <i>useLastModified</i> is checked, the <i>Last-Modified</i> value is re-submitted in an <i>If-Modified-Since</i> request header the next time that resource is retrieved. That then allows the web server to return a <i>304 - Not Modified</i> response if the resource is unchanged. If this is not checked, the resource is downloaded and compared with the previously retrieved content. You would only choose to do this if you found that your web server was returning unreliable results for <i>Last-Modified</i> .

Element	Description
Use SSPETag	<p>This option controls a proprietary mechanism used to identify changes to files retrieved by the GET_FILE service.</p> <p>Without this mechanism, file content is retrieved and compared with the previous version.</p> <p>Use of this mechanism allows the content server to return a <i>304 - Not Modified</i> response and so avoid unnecessary downloads.</p>
Default Filename	Specify the filename to be used for URLs where there is no filename explicitly specified.
Page Extension	<p>The extension that is added to page urls that do not otherwise specify an extension. For example, a typical reference to a document in the dynamic site might look like:</p> <p><code>http://myServer/mySite/Section1/DocumentX</code></p> <p>Where <code>DocumentX</code> is the <i>dDocName</i> of the target item in Content Server. For use on a static site, you must add an extension.</p>
User Agent	Enables you to specify a value for the <i>User-Agent</i> http request header used by Site Studio Publisher when crawling the site.
Friendly Url Parameters	<p>Specify a comma separated list of additional parameter names to honor.</p> <p>Site Studio Publisher supports a few built-in values.</p> <p>Friendly Urls already honor parameters ending in <i>NextRow</i> or <i>_dcPageNum</i> and use the parameter name and value to construct the filename for the crawled page.</p> <p>If a URL Parameter affects the appearance of the page, then you need to capture a different copy of the page for each combination of parameter values.</p>
Dynamic Url Parameters	<p>Specify a comma separated list of additional parameter names to honor.</p> <p>Dynamic Urls using the GET_PAGE service already honor <i>dID</i>, <i>dDocName</i>, <i>RevisionSelectionMethod</i>, and <i>Rendition</i>.</p> <p>If a URL Parameter affects the appearance of the page, then you need to capture a different copy of the page for each combination of parameter values.</p>
Additional Services	<p>Select to allow additional services to be crawled.</p> <p>This element enables you to control the service calls that Site Studio Publisher attempts to crawl.</p> <p>There is built-in support for those services that are expected to generate meaningful static content: <i>SS_GET_PAGE</i>, <i>GET_FILE</i>, and <i>GET_DYNAMIC_CONVERSION</i>.</p>
Soft Error Threshold	The number of soft errors allowed. If the defined number is exceeded, publishing fails.
Hard Error Threshold	The number of hard errors allowed. If the defined number is exceeded, publishing fails.
Delete Threshold	The number of objects that can be missing before the crawl is failed. If this number is exceeded, publishing fails.
Soft Error Codes	<p>Enter a list of codes to specify individual error codes that are treated as soft errors.</p> <p>There are no wildcards allowed in this list, you must list each error code explicitly (separated by a comma).</p>

Element	Description
Hard Error Codes	Enter a list of codes to specify individual error codes that are treated as hard errors. There are no wildcards allowed in this list, you must list each error code explicitly (separated by a comma).
Ignore Error Codes	Enter a list of codes to specify individual error codes that are ignored. These codes will not affect the Site Studio Publisher crawl. There are no wildcards allowed in this list, you must list each error code explicitly (separated by a comma).

A.2.2 Dates

The Dates section is used to specify a range of dates that the task will run.

If you select **Manual**, then the Days section options (see "Days" on page A-6) and the Run Time section options (see "Run Times" on page A-7) are greyed out. This is because selecting **Manual** will mean that the task is a task with will run only on-demand.

Figure A-4 Dates Section

Element	Description
Manual / Range	Select to either have a task that runs only when commanded (manual) or to run at regular intervals at least once within a certain range of dates.
Start Date	The first date of a range that the task will run.
End Date	The final date of a range that the task will be run.

A.2.3 Days

The Days section is used to select if the task will run on certain days of the week, or on certain days of the month.

The options in this section will be available only if **Range** was selected in the Dates section (see "Dates" on page A-6).

Figure A-5 Days Section

Element	Description
Days of Week	Select to run the task on the selected days of the week. The task will run on the days selected that are within the range selected in the Dates section (see "Dates" on page A-6).
Days of Month	Select to run the task on particular days of the month. Days can be listed individually, by a range, or in a combination of both, for example: 1, 2, 7-13, 25. The word <code>last</code> is also a valid entry, to denote the last day of a month. It can be entered by itself or in combination as above.

A.2.4 Run Times

The Run Times section is used to specify how often the task runs each day.

The options in this section will be available only if **Range** was selected in the Dates section (see "Dates" on page A-6).

Figure A-6 Run Times Section

Element	Description
Multiple Times Per Day	Select to have a task that runs either at regular intervals each day.
Once per Day	Select to have a task that runs only once per day. Enter a time in the box.
Hours Minutes	Select how often the task should run between the Start Time and the End Time.
Start Time	The time of day the task will start running.
End Time	The time of day the task will stop running.

A.2.5 Options

The Options section is where you set the priority and log levels of the task. You can also set an email address to receive notification after the task runs.

Figure A-7 Options Section

Element	Description
Priority	Select to mark this task as a priority task. This is used to ensure that the more vital tasks are run at a priority over the other tasks.
Log Level	Select the level of log information to write to the logs. Each item in the drop-down list includes the logging levels above it. For example, selecting INFO includes logs of not only INFO items but also WARN and ERROR .
Email Notification	Enter an email address to send notification of the task completion (or error) to.

A.2.6 Include List

The Include List section is used to specify which parts of the site (by URL) should be included in publishing.

If you list a URL for inclusion in this section, and it appears in a listed FilterSet type 'exclude', then it will not be included in the publishing.

Figure A-8 Include List Section



Element	Description
Hide Include List	Click to minimize the Include List section.
URL	Enter a URL to place on the Include List for Site Studio Publisher to crawl. This can be a regular expression.
Remove	Click to remove the URL from the list.
Add New Item	Click to add another field to enter an additional URL for the Include List.

A.2.7 Exclude List

The Exclude List is used to specify which parts of the site (by URL) should be excluded from publishing.

Figure A-9 Exclude List Section



Element	Description
Hide Exclude List	Click to minimize the Exclude List section.
URL	Enter a URL to place on the Exclude List for Site Studio Publisher to avoid. This can be a regular expression.

Element	Description
Remove	Click to remove the URL from the list.
Add New Item	Click to add another field to enter an additional URL for the Exclude List.

A.2.8 FilterSets

Filtersets are controls used to transform content once it is downloaded from crawling, before it is published. Filters are applied in the order that they appear in the Filterset.

Figure A–10 FilterSets Section

Element	Description
Add New Filterset	Expands to display an additional set of fields to enter information for another filterset.
Move Up Move Down Remove	Select to move the filterset up or down in the list relative to other filtersets, or to remove it completely.
Type	<p>The type attribute specifies the context in which the filterset should be invoked. The types available are:</p> <p><i>transform-content</i>: for URLs that match this filterset, the child filter elements are applied during download, transforming the content in the manner specified by these filters.</p> <p><i>transform-link</i>: for URLs that match this filterset, the child filter elements are applied during download. In this case, the filter elements are only applied to links found in the current downloaded file, not to the entire content.</p> <p><i>exclude</i>: for URLs that would normally be included, match this filterset, and the URL matches, content is not downloaded.</p>

Element	Description
Path	<p>The path attribute is a wildcard pattern to match the file path of the URL (the part following the URL's host name) currently being downloaded.</p> <p>A Java regular expression syntax is used for pattern matching. Use the following reference:</p> <p>http://java.sun.com/j2se/1.4.2/docs/api/java/util/regex/Pattern.html</p>
Hostname	Enter a value to match the URL's host name (the domain name part of the URL) with the specified host name.
Port	Enter a value to match a URL's port number.
Mime Type	Enter a value to match the URL's MIME-type. The value that starts with the value specified by this type is considered a match; therefore, <code>text</code> matches both <code>text/html</code> and <code>text/xml</code> .
Show Filters	<p>Expands to display the filters specified in the FilterSet.</p> <p>When opened, the link becomes Hide Filters, to close the list of individual filters.</p>
Add New Filter	Adds an additional set of fields to enter a new filter.
Move Up Move Down Remove	Select to move the filter up or down in the list relative to other filters, or to remove it completely.
Regex	A java regular expression used to identify text to be replaced in the content of the current file or in the link.
Replacement	The value replacing the text identified by the <i>regex</i> attribute.
Path Regex	A Java regular expression used to identify a path and file to replace in the context of the current file. This value is only meaningful within an enclosing <i>transform-content</i> filterset.
Path Replacement	The replacement value used to change the filename.
Global	Controls the behavior of the substitution. If Global is checked, every match for the <i>regex</i> expression is replaced. If unchecked, only the first occurrence is replaced.
Ignore Case	Controls the behavior of the <i>regex</i> expression. If Ignore Case is selected, then the case of the expression is ignored when comparing the <i>regex</i> .

A.2.9 Triggers

Triggers enable the replication engine to run commands either before a package is downloaded or after the download is complete. The types of triggers available are **cmd**, **http-post**, **http-get**, and **http-soap**.

Figure A-11 Triggers Section

The screenshot shows a 'Triggers' configuration window. It contains the following fields and options:

- Username:** administrator
- Password:** masked with dots
- Authentication:** BasicAuth (dropdown menu)
- Move Up | Move Down | Remove:** text links
- Type:** http-soap (dropdown menu)
- Command:** http://SERVER:8889/axis/services/OfferService
- Command Data:** cns-soap-trigger.xml
- Response File:** cns-soap-response.xml
- SOAPAction:** empty text box
- Run the trigger even if the crawl failed
- Run the trigger if there were no changes
- Ignore trigger failure
- Log the response
- Add New Trigger:** button at the bottom left

Element	Description
Show Triggers	Displays the fields required to enter the most basic information for a trigger.
Add New Trigger	Displays additional information to add multiple types of triggers.
Username	The username used for authentication. You cannot specify different authentications for each trigger. All triggers in a task must have the same credentials and authentication method.
Password	The password used for authentication. You cannot specify different authentications for each trigger. All triggers in a task must have the same credentials and authentication method.
Authentication	Select to authenticate the user either with basic authentication or with a custom form.
Move Up Move Down Remove	Select to move the trigger up or down in the list relative to other triggers, or to remove it completely.
Type	Specifies the type of trigger.
Command	Command line arguments to pass through the trigger. Only used if the type is set to cmd .
Command Data	The path, relative to the task's output path, of the file to be uploaded. Only used if the type is set to http-post or http-soap .
Response File	Specifies the full path to a file to capture the response.
SOAPAction	Sets the value for the SOAPAction HTTP request header field. Only used if the type is set to http-soap .
Run the trigger even if the crawl failed	Specifies whether the command should be run, even if the job encountered errors.
Run the trigger if there were no changes	If checked, the trigger will run even if there has been no change in content.
Ignore trigger failure	Specifies whether further trigger execution will happen. If checked, further triggers will run, even if the current trigger fails.
Log the response	Specifies whether to write the response to the log.

A.3 Task Info

The Task Info screen is used to view information about the most recent task initiating the specified Site Studio Publisher crawl.

Figure A–12 Task Info screen

Task Info	
Administration --> Site Studio Administration --> Site Studio Publisher --> Task Info	
<hr/>	
Job Summary	
Task ID	2
Description	runondemand
Source Url	http://MYVMWARE/stellent/idcplg?IdcService=SS_GET_SITE_PUBLISH_REPORT&siteId=Loafers
Start Time	November 16, 2009 2:42 PM
End Time	November 16, 2009 2:43 PM
Status	Completed
<hr/>	
Links Summary	
Total	76
Modified	0
Added	75
Deleted	0
Examined	77
Redirected	0
Downloaded	75
Not Found	2
<hr/>	
Error Summary	
Hard error count	1 (99)
Soft error count	1 (99)
<hr/>	
View Log	

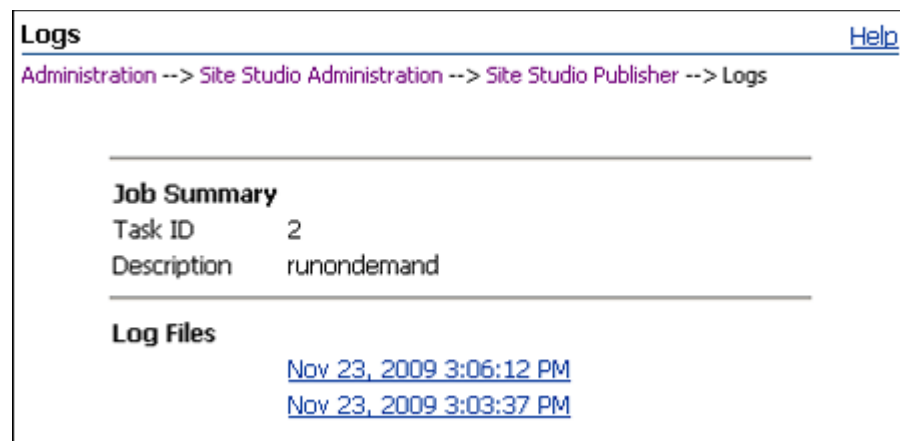
Element	Description
Task ID	The ID number of the task.
Description	The description of the task.
Source Url	The source URL of the site being published.
Start Time	The time the task started.
End Time	The time the task completed.
Status	The completion status of the task.
Total	Total number of URLs (including files such as CSS, JavaScript, and so forth) that make up the crawled site.
Modified	Number of URLs that were modified since the most recent crawl of the site.
Added	Number of URLs added since the most recent crawl of the site.
Deleted	Number of URLs deleted since the most recent crawl of the site.
Examined	The number of URLs followed while crawling the site.
Redirected	The number of URLs redirected while crawling the site.
Downloaded	The number of examined links that were actually downloaded.

Element	Description
Not found	The number of links followed that received an error.
Hard Error Count	Total number of "hard" errors found. The number in parenthesis is the maximum allowed.
Soft Error Count	Total number of "soft" errors found. The number in parenthesis is the maximum allowed.
View Log	Opens the log file for the task in the View Log screen (see " View Log " on page A-13).

A.4 Logs

Lists the job summary and different log files generated for the specific task.

Figure A-13 Log Files



Element	Description
Task ID	The ID number of the task.
Description	The description of the task, as entered in the Publisher Settings section on the Add/Edit Task page (see " Publisher Settings " on page A-2).
Log Files	A link for each date and time that the task was run. Each link opens the log file for that specific run on the View Log page (see " View Log " on page A-13). If the log file is very long, it will be split across multiple files, and each file will have a link with the same timestamp.

A.5 View Log

The View Log screen is where the log file can be viewed. When the log file is long, the **Previous** and **Next** buttons are available to page through the file.

Figure A-14 View Log Screen

```

View Log
Administration -> Site Studio Administration -> Site Studio Publisher -> Logs -> View Log
[Previous] [Next]
INFO 2009-11-25 09:24:44,581 - ----- Running JSP
INFO 2009-11-25 09:24:44,581 - locale: file = C:\oracle\jdev\workspace\data\SiteStudio\Publisher\output\SITE
INFO 2009-11-25 09:24:44,581 - archName: English
INFO 2009-11-25 09:24:44,581 - doCacheControl: false
INFO 2009-11-25 09:24:44,581 - doPublishNew: false
INFO 2009-11-25 09:24:44,581 - doForceAnalyze: false
INFO 2009-11-25 09:24:44,581 - doForceDownload: false
INFO 2009-11-25 09:24:44,581 - testWCAsWebServices: true
INFO 2009-11-25 09:24:44,581 - treatNonPageErrorsAsCritical: true
INFO 2009-11-25 09:24:44,581 - treatManifestErrorsAsCritical: true
INFO 2009-11-25 09:24:44,581 - ----- Connecting: http://myvmware/stellent/ldplog?IdcService=SS_SET_SITE_PUBLISH_REPOSTateId=Monday
INFO 2009-11-25 09:24:45,011 - Response: 200 OK
INFO 2009-11-25 09:24:45,011 - opened in 8:00:00.138
INFO 2009-11-25 09:24:45,041 - Manifest Version: 7.7.0.1
INFO 2009-11-25 09:24:45,041 - adding 1 style from manifest
INFO 2009-11-25 09:24:45,041 - added Root Link: http://MYVMWARE/Monday/inSto.htm
INFO 2009-11-25 09:24:45,041 - adding 1 fragment assets from manifest
INFO 2009-11-25 09:24:45,041 - added Root Link: http://MYVMWARE/stellent/fragments/dynamicListGeneric/DynamicListGeneric.css
INFO 2009-11-25 09:24:45,041 - adding 0 declared assets from manifest
INFO 2009-11-25 09:24:45,041 - adding 0 site assets from manifest
INFO 2009-11-25 09:24:45,041 - adding 0 custom property assets from manifest
INFO 2009-11-25 09:24:45,041 - Site URL = http://MYVMWARE/stellent/ldplog?IdcService=SS_SET_SITE_PUBLISH_REPOSTateId=Monday
INFO 2009-11-25 09:24:45,041 - -----

```

Element	Description
Previous	Takes you to the previous page in the log.
Next	Takes you to the next page in the log.

Index

A

absolute URL, 1-2
Add New Task screen, A-2
 Dates section, 2-5, A-6
 Days section, 2-6, A-6
 Exclude list section, 2-7, A-8
 FilterSets section, 2-7, A-9
 Include List section, 2-7, A-8
 Options section, 2-6, A-7
 Publisher Settings section, 2-2, A-2
 Run Times section, 2-6, A-7
 Triggers section, 2-8, A-10
appended parameters, 1-3

C

cmd trigger, 2-9
Content Server environment, 1-2
Contributor
 Publish Now feature, 2-2

D

Dates section, 2-5, A-6
Days section, 2-6, A-6

E

Edit Task screen, A-2
 Dates section, 2-5, A-6
 Days section, 2-6, A-6
 Exclude list section, 2-7, A-8
 FilterSets section, 2-7, A-9
 Include List section, 2-7, A-8
 Options section, 2-6, A-7
 Publisher Settings section, 2-2, A-2
 Run Times section, 2-6, A-7
 Triggers section, 2-8, A-10
editing a task
 scheduling, 2-5, 2-6
Exclude List section, 2-7, A-8

F

filter
 re-applying, 2-3

filter application, 2-7
filter element, 2-7
filters, 2-7, A-9
Filters section, 2-7, A-9
filterset
 evaluation, 2-7
FilterSets section, 2-7, A-9

H

HTML
 handling in Publisher, 1-2
http-get, 2-9
http-post, 2-9
http-soap, 2-9

I

Include List section, 2-7, A-8
installation, 1-2

J

JavaScript
 general patterns, 1-3
 generated by Site Studio, 1-3
 handling in Publisher, 1-3
 parameters, 1-4
 patterns used by Publisher, 1-3
 Site Studio-specific patterns, 1-3, 1-4

L

Logs screen, A-13

M

minimum requirements, 1-4

O

Options section, 2-6, A-7

P

properties
 URL Directory Name, 1-4

URL Page Name, 1-4
Publisher control
 Publish Now, 2-2
 Publish This Page, 2-2
Publisher Settings section, 2-2, A-2
Publisher Tasks screen, A-1

R

relative URL, 1-2, 1-3
Run Times section, 2-6, A-7

S

screen
 Add New Task, A-2
 Edit Task, A-2
 Publisher Tasks, A-1
Site Studio Contributor
 Publish Now feature, 2-2
SOAP, 2-9
system requirements, 1-4

T

Task Info screen, A-12
task scheduling, 2-5, 2-6
trigger, 2-8
 cmd, 2-9
 http-get, 2-9
 http-post, 2-9
 http-soap, 2-9
Triggers section, 2-8, A-10

U

URL
 absolute, 1-2
 relative, 1-2, 1-3
URL Directory Name property, 1-4
URL Page Name property, 1-4

V

View Log screen, A-13