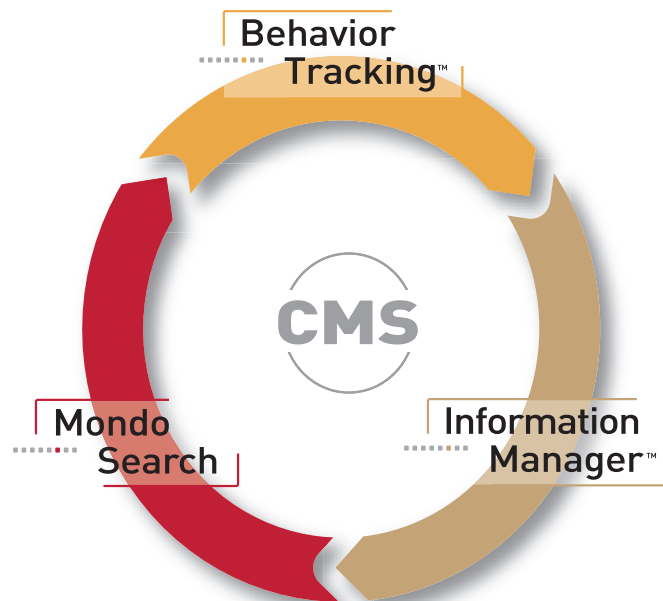




# MondoSearch for EPiServer 2.0

– User's Guide



---

## **Important Notice**

### **Disclaimer of Warranty**

Mondosoft A/S makes no representations or warranties, expressed or implied, by or with respect to anything in this document, and shall not be liable for implied warranties of merchantability or fitness for a particular purpose or for any indirect, special, or consequential damages.

### **Copyright Notice**

Copyright © 2006 Mondosoft A/S. All rights reserved. The copyright of this document is vested in Mondosoft A/S.

This document may only be reproduced in whole or in part, stored in a retrieval system, transmitted in any form, or by any means electronic, mechanical, photocopying, or otherwise, with the prior permission of Mondosoft A/S.

### **Trademarks**

MondoSearch is a registered trademark of Mondosoft A/S. BehaviorTracking, InformationManager, SearchNames, BehaviorMatch and SearchHeaders are trademarks of Mondosoft A/S.

Microsoft is a registered trademark and Windows is a registered trademark of the Microsoft Corporation.

**Press date:** October 9, 2006

---

# Table of Contents

<b>Chapter 1: Introduction to MondoSearch for EPiServer</b>	<b>7</b>
Benefits of Integration . . . . .	8
Features Added to EPiServer . . . . .	9
Integration Points . . . . .	10
<i>Web-Service Based.</i> . . . .	10
<i>The Search/Result Page</i> . . . . .	10
<i>Expanded Metadata</i> . . . . .	10
<i>Security Model</i> . . . . .	11
<i>Crawler Control and Scheduling</i> . . . . .	11
<i>The Click Page</i> . . . . .	11
Categorized Search Results . . . . .	11
Detailed Logging and Reporting with BehaviorTracking . . . . .	13
<i>Why Track User Behavior?</i> . . . . .	13
<i>How are Statistics Gathered?</i> . . . . .	18
Targeted Promotions with SearchHeaders . . . . .	19
Go Straight to the Best Content with SearchNames. . . . .	20
 <b>Chapter 2: Installing MondoSearch for EPiServer</b>	 <b>21</b>
Before You Install. . . . .	22
Installing the Integration Package Files . . . . .	22
<i>Installing on the Standard EPiServer Demo Site</i> . . . . .	22
<i>Installing on Your Own Custom Site</i> . . . . .	23
Required Initial Configuration Settings . . . . .	24
<i>Configuring MondoSearch.</i> . . . .	24
<i>Configuring BehaviorTracking.</i> . . . .	26
<i>Configuring InformationManager</i> . . . . .	26
<i>Configuring EPiServer.</i> . . . .	27
Configuring Crawler Start Points and Log In. . . . .	31
<i>Crawler Settings on the InSite</i> . . . . .	31
<i>Modify the Robot Commands on Your Login Page.</i> . . . .	35
Testing the Installation . . . . .	35
Blocking Unneeded CGI Access to MondoSearch . . . . .	35
 <b>Chapter 3: The META-Tag Emitter</b>	 <b>37</b>
Purpose of the META-tag Emitter . . . . .	38
META-Tags Added by the Emitter . . . . .	38
Placing the META-Tag Emitter on Your Pages . . . . .	39
Controlling Access to the Metadata . . . . .	39

<b>Chapter 4: Optimizing Your Site for Search</b>	<b>41</b>
Optimizing Page Titles and Descriptions . . . . .	42
Improving Page Rankings . . . . .	42
Marking Content for Exclusion from the Search Database . . . . .	43
<i>Excluding Entire Documents</i> . . . . .	43
<i>Excluding Parts of Documents</i> . . . . .	44
<b>Chapter 5: MondoSearch Administration</b>	<b>45</b>
Accessing the MondoSearch Administration Settings . . . . .	46
“General” Settings . . . . .	46
<i>Controlling Edit-Mode Features’ Appearance and Availability</i> . . . . .	46
<i>Setting the BehaviorTracking Date Span</i> . . . . .	48
Configuring Web Service Communications . . . . .	49
<i>Web Service URLs</i> . . . . .	50
<i>MondoSearch License Code</i> . . . . .	50
<i>Connection Strings for the Integration Services</i> . . . . .	51
<i>Demo Settings</i> . . . . .	53
Scheduling the Crawler . . . . .	53
<i>Setting the Crawler Schedule and E-Mail Alerts</i> . . . . .	53
<i>Triggering the Crawl Schedule with the Background Job</i> . . . . .	55
Setting the Crawler Agent ID . . . . .	55
Using the Diagnostics Tool . . . . .	56
<b>Chapter 6: Manual Crawler Control and Monitoring</b>	<b>59</b>
Accessing the MondoSearch Control Center . . . . .	60
Crawler Monitoring and Status . . . . .	61
Manual Crawler Control and Publishing . . . . .	62
Reading the Crawler Log . . . . .	64
<b>Chapter 7: Search/Result Page Configuration in EPiServer</b>	<b>65</b>
Configuring Search Options . . . . .	66
Graphic Design and Advanced Customization . . . . .	69
<b>Chapter 8: Setting up MondoSearch Categories</b>	<b>71</b>
Planning Your Categories . . . . .	72
How MondoSearch Categorization Works . . . . .	72
Making Category Assignments . . . . .	73
Categorizing Binary Document Types . . . . .	75

---

<b>Chapter 9: Providing Authorized Result Lists</b>	<b>77</b>
Introduction . . . . .	78
Configuring the Authentication System . . . . .	80
<b>Chapter 10: Editor Search</b>	<b>81</b>
Editor Search Appearance and Availability . . . . .	82
Using the Editor Search . . . . .	82
<b>Chapter 11: Working with BehaviorTracking in EPiServer</b>	<b>85</b>
Finding BehaviorTracking in EPiServer . . . . .	86
The BehaviorTracking Date Span Setting . . . . .	87
BehaviorTracking Overview . . . . .	88
<i>Overview—Searches per Day</i> . . . . .	88
<i>Performance Map</i> . . . . .	89
Searches & Clicks . . . . .	92
<i>Searches &amp; Clicks Overview</i> . . . . .	92
<i>Top Search Terms</i> . . . . .	93
<i>Top Not-Found</i> . . . . .	94
<i>Top Not-Clicked Terms</i> . . . . .	95
<i>Top Links Chosen</i> . . . . .	97
Search Sessions . . . . .	98
<i>Search Sessions Overview</i> . . . . .	98
<i>Most-Recent Sessions</i> . . . . .	99
<i>Top Visitor Domains</i> . . . . .	100
Trends . . . . .	101
<i>Search-Term Trends</i> . . . . .	101
<i>BehaviorMatch</i> . . . . .	102
Term Details . . . . .	103
<i>Related Terms</i> . . . . .	103
<i>Most-Recent Sessions that Include this Term</i> . . . . .	104
<i>Pages Chosen</i> . . . . .	104
Session Details . . . . .	105
Page Details . . . . .	106
<i>Viewing Details for a Selected Page</i> . . . . .	106
<i>Interpreting the Page Details</i> . . . . .	107
Acting on BehaviorTracking Data . . . . .	108
<b>Chapter 12: InformationManager Features in EPiServer</b>	<b>113</b>
Working with SearchHeaders . . . . .	114
<i>Creating SearchHeaders</i> . . . . .	114
<i>Editing SearchHeaders</i> . . . . .	116

<i>Deleting SearchHeaders</i> .....	116
<i>Dynamic SearchHeaders</i> .....	116
SearchNames .....	116
<i>Creating SearchNames</i> .....	117
<i>Editing SearchNames</i> .....	118
<i>Deleting SearchNames</i> .....	118
<i>Dynamic SearchNames</i> .....	118
 <b>Appendix A: Using the Native MondoSearch Interfaces</b>	 <b>121</b>
Search Configuration in the MondoSearch InSite .....	122
Using the Stand-Alone BehaviorTracking Application .....	123
Using the Stand-Alone InformationManager Application .....	123
 <b>Appendix B: Advanced Integration and Development</b>	 <b>125</b>
Setting Custom Search Options with MQL .....	126
Building Your Own Search/Result Pages .....	126
Custom Integration and Feature Development .....	127
 <b>Appendix C: Installing Mondosoft Integration Services</b>	 <b>129</b>
Requirements .....	130
Package Contents .....	130
Installing the Files .....	131
Configuring the Installation for Use .....	132
<i>Specify the Database Connection String for BehaviorTracking</i> .....	132
<i>Create an Information Manager User if Needed</i> .....	133
 <b>Contacting Mondosoft</b>	 <b>135</b>



# CHAPTER 1

## Introduction to MondoSearch for EpiServer



This integration package makes it easy to add the features of MondoSearch™, InformationManager™ and BehaviorTracking™ to sites managed with EpiServer™. The result is a vastly improved search feature that makes users more productive and gives managers more insight into how the site and its search engine are used and how they can be improved.

## Benefits of Integration

EPiServer helps your organization to produce a rich and attractive Web site that is easy to author, update and manage. However, the standard search engine delivered with the system provides only very basic search functionality. Integration with MondoSearch applications provides a vastly improved search feature and many related monitoring, analysis and application features that will make your site even more valuable to your organization and its users. Benefits include:

- *Improve the productivity benefit of your Web site*  
MondoSearch introduces many features that help make users more productive by directing them to the exact information they need as quickly as possible. Users do not need to learn new skills or habits to be more productive with MondoSearch; instead the search engine adapts to its users, reacting the way most people would naturally expect. Features such as synonyms, phonetics and stemming help ensure that results include all relevant documents, while its categorization and advanced sorting algorithms help ensure that the best documents are presented right up front. For problematic searches, MondoSearch provides additional help in the search-again form that will help users refine their queries for better results the second time.
- *Create new, innovative site features that incorporate search and BehaviorTracking data*  
Because the MondoSearch integration is based on .NET Web services, your developers can integrate search and BehaviorTracking data throughout the site to create entirely new types of features. One example of this is provided in the default installation package; right on the welcome page of your site, users will be able to view and select links to the pages most often selected after using the search engine.
- *Improve site content and search response based on user behavior*  
With BehaviorTracking, MondoSearch learns from the way real people actually use search on your site and gives feedback to site managers that will help them tune the search engine settings and site content.
- *Apply what you have learned*  
With features supplied by Mondosoft InformationManager, you will be better able to take advantage of the information supplied by BehaviorTracking. For example, you can create search-word synonyms that will direct users to documents that describe concepts related to the original search even though those documents use alternative terms (such as “film” vs. “movie”) or describe products that compete with commonly used trademarks (such as Walkman™, iPod™, Frisbee™ or Band-Aid™). BehaviorTracking reports often-used search words that produce no results—these are often good contenders for synonyms or META-tags.
- *Faster editing (search for template)*  
Authors and editors of your EPiServer site can also take advantage of the benefits offered by MondoSearch as they work. The search engine is available in EPiServer Edit Mode, making it easier to locate existing content pages that need to be edited or updated.



## Features Added to EPiServer

After integration, all of the many features of MondoSearch, BehaviorTracking and InformationManager become available to your site. Many of the most-used features become available directly within the EPiServer authoring and/or browsing environment, including:

- *Categorized search results*  
Users find what they need much faster when they can break results down into categories.
- *MondoSearch advanced ranking and sorting*  
MondoSearch considers all of the available information to help make sure that it puts the most relevant documents at the very top of its result list, even when users supply just a one or two search words as input. Intelligent processing of multi-word queries plus context-based evaluation of found words helps MondoSearch bring the best results forward. For example, MondoSearch does not simply return an arbitrary list of documents that contain a submitted search word; instead it ranks each document based on whether that word appears many times on the page, appears in headlines, appears in keyword META-tags and other factors.
- *Integrated Crawler control and monitoring*  
Site administrators can start, stop and monitor the MondoSearch Crawler from within the EPiServer environment.
- *Authorized search results*  
Make sure that search results list only those documents that the current user has permission to view. This makes the search results both more secure and more user friendly.
- *Find HTML, PDF, Office and Flash documents*  
MondoSearch not only indexes your HTML pages, but also the full content of your Adobe Acrobat (PDF), Microsoft Office (Word, Excel and PowerPoint) and Adobe Flash documents. All of these document types will be included in search results.
- *Detailed multi-Language support*  
MondoSearch automatically identifies a wide selection of languages and provides full support for Unicode, which enables two-byte languages (such as Arabic and Japanese) to be indexed and searched. Result lists indicate the language of each page and users can filter results by language. Many advanced features, including stemming and stop-words, adapt to the language of each individual page.
- *Fully customizable search form and result presentation*  
Because the MondoSearch integration is based on .NET Web services, your site developers can create totally customized search and result pages. A full-featured search/result page is provided with the package to help you get started right away.
- *BehaviorTracking logging and analysis tools*  
Keeps a detailed record of how people use the search engine, including what they searched for, how they modified their queries and what pages they selected afterwards. Analysis tools are integrated right into the EPiServer edit-mode interface, so managers and page authors can make use of this data as they add content and improve the site.

- *Key-word suggestions based on user behavior*  
BehaviorTracking is able to deduce, based on the record of how users refine their queries, which search words users seem to expect to find on each page. In EPiServer edit-mode, these likely keywords are listed for each page so page authors can choose to add them with a click of the mouse.
- *Most-wanted links on home page*  
As an example of the additional power made available to site developers when they can access search and BehaviorTracking data from any page, the integration package enhances the EPiServer welcome page to include a display of the links most selected by search-engine users.

## Integration Points

### Web-Service Based

All of the real-time integration features between EPiServer and the MondoSearch applications are implemented using .NET Web services, which are called using standard .NET techniques. This makes the solution very open, flexible and standards-compliant. For example, when a user makes a search, EPiServer simply forwards the query to the MondoSearch SearchService, which processes it and returns a data set back to EPiServer. The search-result page then formats and displays the data set using standard .NET techniques.

### The Search/Result Page

The integration package includes a .NET page pre-programmed to communicate with the Mondosoft Web Services and to display result sets. The page is prepared to respond to standard configuration settings, including Web service addresses and search options, which can easily be made using new settings added to the EPiServer Administration Mode interface.

The search/result page is a standard .NET (aspx) document, which experienced developers can easily open and customize using a tool such as Microsoft Visual Studio. Developers can make basic modifications by editing the default page or can even begin an entirely new project that incorporates calls to the various MondoSearch applications and combines the returned data with local data in new and innovative ways.

### Expanded Metadata

Included in the integration package is a new control that generates META-tags to support several advanced MondoSearch features, such as categorization and result-list authorization. This control should be included on every page served by EPiServer. In the default installation, this is handled automatically by placing the control in the header template of EPiServer. The control is normally configured to generate its data only for the MondoSearch Crawler program; other crawlers and standard Web browsers will not receive the extra data.

## Security Model

BehaviorTracking displays, InformationManager controls and MondoSearch administrations features are all integrated into the standard EPiServer editing and administration interface. All of these displays and settings are therefore protected behind the standard EPiServer user authentication and security systems. Site administrators can furthermore restrict access to the various integration features within the edit interface as needed.

The various MondoSearch web services only respond to requests that include valid and encrypted authentication details, which are configured within EPiServer.

For sites that include both public and private areas, MondoSearch delivers result lists that include only those documents that the searching user has permission to read. Though EPiServer would refuse to serve protected documents to unauthorized users anyway, keeping their titles and descriptions out of search results preserves security and provides a better service to users.

## Crawler Control and Scheduling

MondoSearch populates its search database using a Crawler program, which loads each page of your site, indexes all of the text and metadata it finds and follows all links until the entire site has been grabbed. MondoSearch for EPiServer adds interfaces both for initiating manual Crawler runs and for establishing a regular crawl schedule. These controls are available in the standard administration interface of EPiServer.

## The Click Page

To enable the solution to generate a complete search log (for BehaviorTracking) and to recreate frames correctly, all searches and result-link selections must be processed by a page called the *click page*. Each time a user searches or clicks on a result link, the search/result page actually targets the click page, which logs the search action and then redirects the request as needed (searches to MondoSearch and clicks to EPiServer).

Users should never access the click page directly. Therefore, it is installed by the integration package in a way that makes it invisible in the site structure. The supplied search/result page is pre-coded to use the standard click page at its default location.

## Categorized Search Results

Categorization is one of the most important and unique features of MondoSearch. It enables search results to be presented in a way that makes it easy for users to find what they are seeking very quickly.

When users view search-results, the first result page displays an overview that shows the best (e.g. top 3 or 5) matches from each of the relevant categories, as shown below in Figure 1. The user can load a result page direct from here, or can instead choose to see all the matching pages from a single category by clicking on the “show all” link for that category.

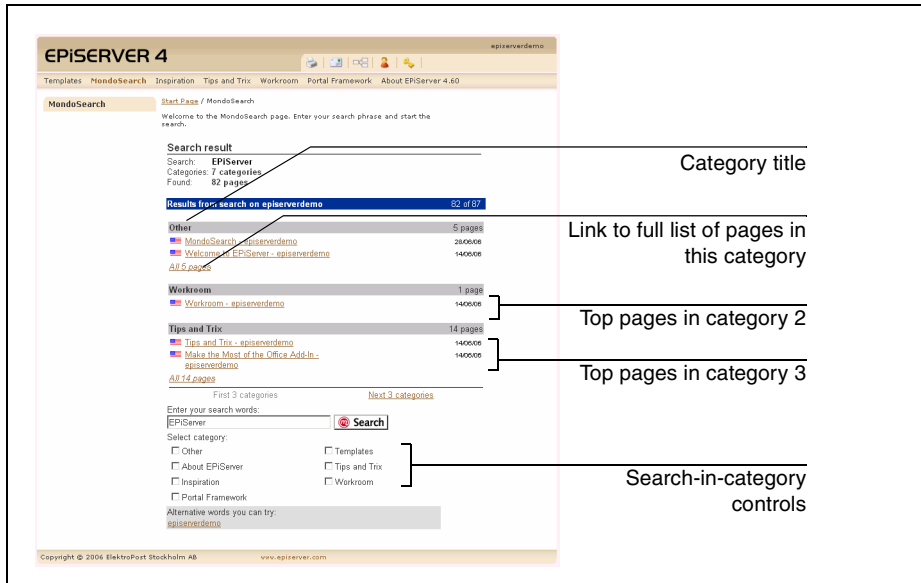


Figure 1 *Categorized search results.*

Page authors and editors specify the category for each page by editing its “dynamic properties” in EPiServer Edit Mode. Category settings can be inherited, which makes it easy to apply a single category to all of the pages in a given branch of your site structure. Instructions for making this setting are given later in this chapter.

**Note:** EPiServer also includes its own system for applying categories to pages. This is not related to MondoSearch categories—the two systems are totally independent and do not affect each other.

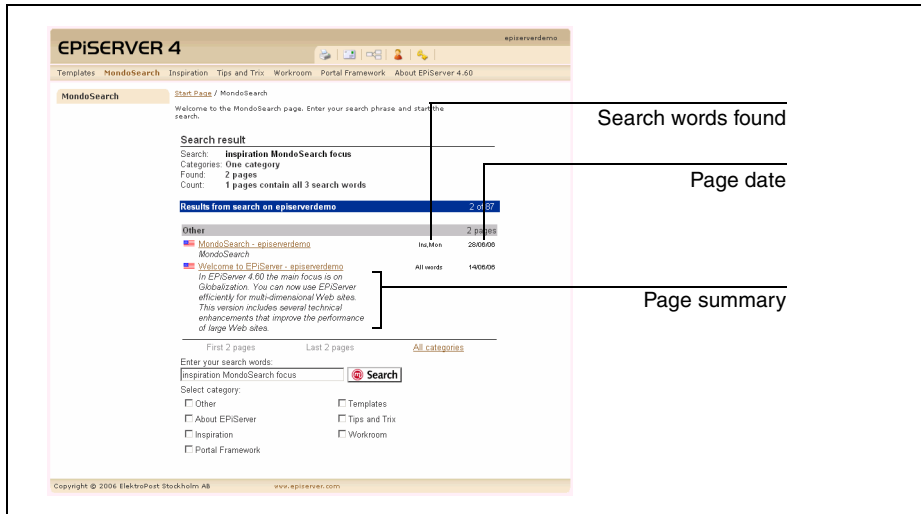


Figure 2 Single-category search results. This page appears after a user clicks on a “show all” link on the multi-category page shown in Figure 1, or if a user searches in a single category only.

## Detailed Logging and Reporting with BehaviorTracking

BehaviorTracking provides advanced search-log analysis tools that enable authors, designers and managers to evaluate how well the search engine is meeting the needs of its users, to see what users are looking for and to decide how best to improve the site content and search engine settings. Its features are among the most important of the new functionality added to EpiServer by the integration package.

### Why Track User Behavior?

#### Your Site's Most Popular Feature

On many Web sites, the most-used feature is the search engine. But the search engine is not only a user's most direct route to the information he or she is seeking, it is also an interactive tool, which means that visitors must provide information in order to receive it.

The search engine therefore knows what people asked for, what they got back and where they went after getting it. This tells more than a simple log of where people happened to wander on the site; it tells us what they wanted to find; it tells us why they came. And it doesn't ask them afterwards—it catches them in the act.

To make this information available to you, BehaviorTracking combines a powerful logging feature with an online statistics-viewing package.

## **Tracking User Interests**

BehaviorTracking stores information about every session that it holds with each user. Each time your search engine receives a query, BehaviorTracking notes:

- What search words were used
- What pages were displayed
- Which of the found links were followed
- The date and time of each session
- Where the request came from (the IP address of the user's router)
- And more...

## A Quick Overview

As you can imagine, after several days of use by hundreds or even thousands of visitors, the volume of logged data can become enormous. To help you identify trends and extract useful information from this vast sea of data, BehaviorTracking provides a set of powerful statistical-analysis tools.

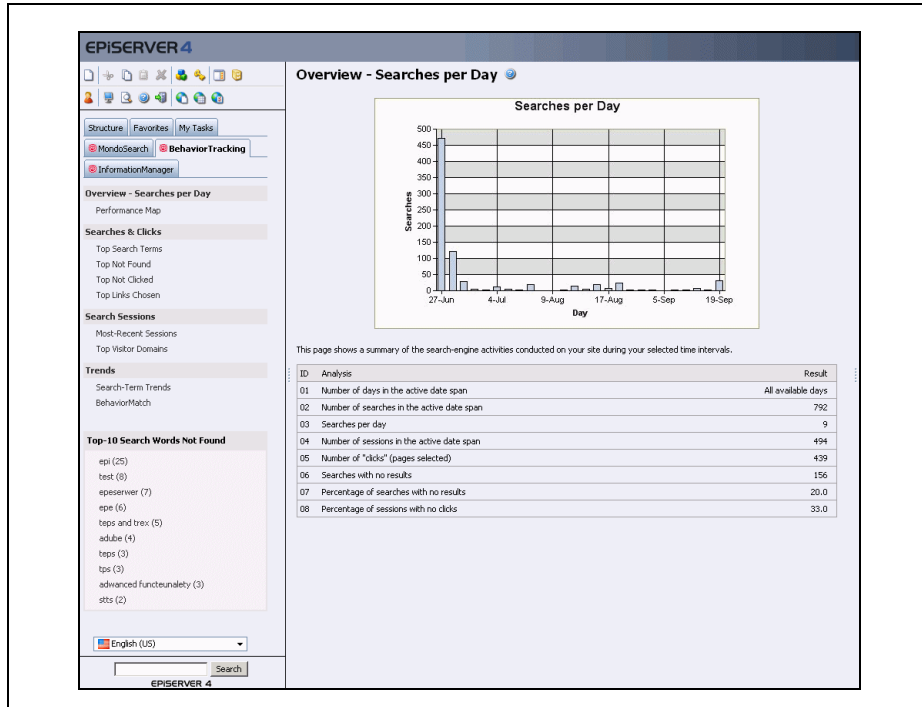


Figure 3: The overview, shown here, gives a summary of all the most vital statistics. Many more details are available by exploring the menu items shown on the left.

## Important Questions

After building up data from several days of logging, we can extract useful statistics to answer all of the following questions (and more):

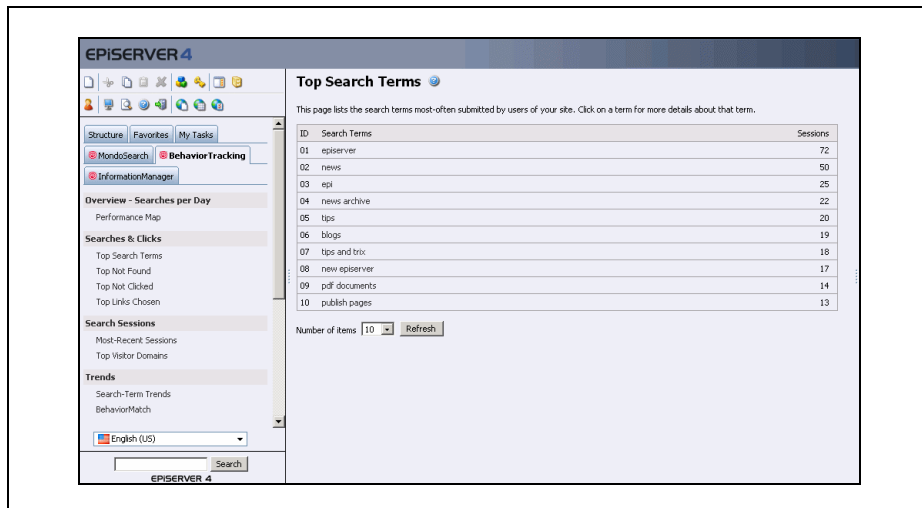
- What search terms were the most asked-for?
- Which popular search terms produced no results?
- Which pages were most-often loaded after a given search term?
- Which pages were most-often loaded after search overall?
- Which terms were used to find any given page?
- How many sessions failed to result in a followed link (the user did not find what he or she was looking for)?
- When is the search engine most used?
- Who is using the search engine the most?

The answers to these questions will interest both people working directly with the site and others working at the organization that is running it.

## Improve Your Search Feature

The information delivered by BehaviorTracking will help webmasters organize the site, choose helpful META-tags and fine-tune their search-engine settings to help optimize search results.

For example, by analyzing the list of most-used search terms, webmasters can look for words that were often misspelled and add these to the keyword META-tags of pages containing the information that users were obviously looking for.



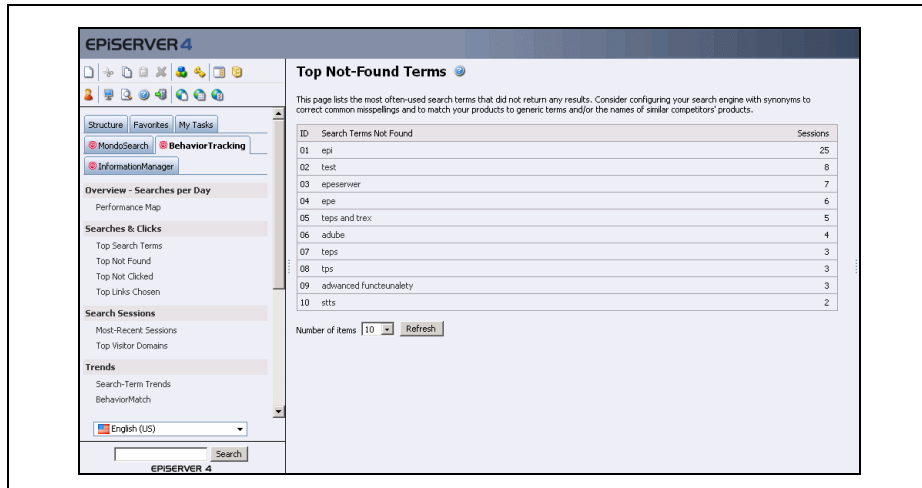
*Figure 4: View the terms most often searched for. Click on a term to view the pages most-often loaded afterwards.*



## Improve Customer Relations

BehaviorTracking will help people from the Marketing department to design new campaigns and judge the success of current campaigns.

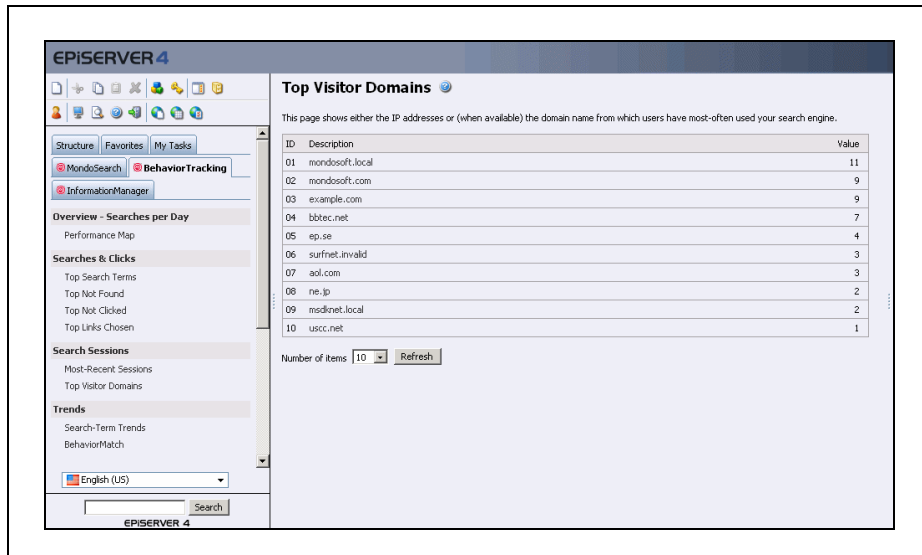
You can see when visitors are often searching for a product that you do not currently carry and possibly add it to your inventory. You can also see when users are looking for a discontinued product and direct them to an updated item.



*Figure 5: Check the list of unsuccessful search words to find common misspellings, discontinued products, competitors' products, and possibilities for new campaigns.*

## The Business Advantage

Because some of your site's visitors are sure to be competitors, the IP-address identifiers will help business people spot “unusual” interest in their company or some certain product.



*Figure 6: Find out who has been using your search engine, how long they used it and what they were searching for.*

## How are Statistics Gathered?

### Tracking Sessions

BehaviorTracking considers all search requests coming from a single IP address with no more than ten minutes break between requests to belong a single session. You should consider a session to represent a full attempt by a user to find a specific piece of information.

### Counting Searches

BehaviorTracking counts each unique search term that occurs in a single session as a single search, regardless of how many times that term is re-submitted during the session. Therefore, actions such as clicking on a paging button to step through multi-page search results and clicking on a “show all results in category” link are counted as being part of the same single “search”. Note, however, that search terms are sensitive to word order, so “red apple” would be considered a different search from “apple red”, even when occurring in the same session.

## Clicks (Chosen Links)

BehaviorTracking takes note each time a user clicks on a link listed on the search results page. We generally consider this to be a successful result—especially when it occurs as the last event in a session. The log stores the page selected, the search term used to find it and other details.

## Targeted Promotions with SearchHeaders

A SearchHeader is a block of HTML that is placed at the top of the search-result page whenever a user submits a particular search word or query. This feature is especially useful for running marketing campaigns with banner announcements that are triggered, for example, whenever a user searches for one of the products or product categories in the campaign. If a user submits a multi-word query, then multiple SearchHeaders can appear if they are defined.

You define a SearchHeader by matching a trigger word to the block of HTML that should be placed on the result page. Your HTML can contain any combination of text and tags, so you can apply text formatting and include graphics and plug-ins, etc. Because Search-Headers can be made out of any type of HTML code, you might also use JavaScript to design them as pop-up windows rather than in-line headers.



*Figure 7: A sample result page that includes a SearchHeader. SearchHeaders enable you to add eye-catching promotional banners in response to specific search words and/or queries.*

The SearchHeaders feature is provided by Mondosoft InformationManager. The integration packages adds controls for creating and managing them to the EPiServer Edit Mode

interface, making it easy for site editors and managers to work with them from within the familiar EPiServer environment.

## **Go Straight to the Best Content with SearchNames**

Sometimes, you may want to make sure that users find a single specific page whenever they submit a particular search word. For example, a search for a particular online game on the site could go straight to the game itself without even showing a result list.

The InformationManager SearchNames feature enables you to define any number of specific queries and match each of them to one or more specific URLs, which will load instead of the usual result page whenever a user submits the associated query.



# CHAPTER 2

## Installing MondoSearch for EPiServer



This chapter explains how to prepare your servers and install the files included with the MondoSearch for EPiServer integration package. It also explains how to make several types of initial configuration settings that will enable the system to function correctly. Finally, we give a little advice for how to optimize your site for search moving forward.

## Before You Install

Before you begin to install the integration package, do the following:

1. If you are using a hosted MondoSearch service, then get the required URLs, connection strings and passwords from your service provider and skip the rest of this step. However, if you are installing MondoSearch locally, then install all of the following:
  - MondoSearch
  - BehaviorTracking
  - InformationManager
  - Integration Services (*see also the note, below*)

See the documentation included with each of the above products for complete installation instructions.

**NOTE:** a limited version of the Integration Services product is included with the MondoSearch for EPiServer integration package. This limited version does not include documentation, but does provide all of the functionality required for the default integration scenario as described in this manual. If you do not have the full version, then please see Appendix C: “Installing Mondosoft Integration Services” for details about how to install the version included with this package.

**IMPORTANT:** Because these are complex server products, you cannot simply rely on the setup wizards to complete the installation for you; be sure to read the documentation and make all installations and configuration settings described for each product.

2. If you did not already do so during the previous step, then run the MondoSearch setup wizard to establish the initial user account and a few essential options. See the *MondoSearch User’s Guide* for instructions.

## Installing the Integration Package Files

MondoSearch for EPiServer includes all of the software required to add MondoSearch features to your EPiServer site. It also includes a few extra (and optional) files that will fine-tune the standard EPiServer demo site.

### Installing on the Standard EPiServer Demo Site

If you would like to learn about the integration package before you install it onto your own site, then you can set up the standard demo site included with EPiServer and install the MondoSearch integration package there. The integration package even includes a few files that will help optimize the demo site for MondoSearch.

1. Make a backup copy of the following files from your EPiServer demo site:
  - `<EPiServerSiteRoot>\default.aspx`
  - `<EPiServerSiteRoot>\templates\units\header.ascx`
  - `<EPiServerSiteRoot>\templates\units\header.ascx.cs`

Where `<EPiServerSiteRoot>` is the root content folder for the demo site.

2. Extract the MondoSearch for EPiServer package and locate the following folder:

---

```
<IntegrationPackageRoot>\SamplePack\webroot_460_and_461
```

---

Where `<IntegrationPackageRoot>` is the folder to which you unzipped the package. Copy the entire contents of this folder to your EPiServer site root and confirm the overwrite of all folders and files.

3. If you are running EPiServer 4.60 then delete the following file (which holds the language cache):

---

```
<EPiServerSiteRoot>\util\temp\<SiteName>.cache
```

---

Where `<SiteName>` is an (often long) name generated by EPiServer. Among other things, the name incorporates the domain name of your site.

(This step enables the language part of the package to be available and preserves the look of the edit/admin GUI in EPiServer 4.60. The problem that makes this step necessary has been fixed in EPiServer 4.61.)

4. If possible, open the demo site project in Microsoft Visual Studio and recompile the entire project.

(This step enables the improvements made by the new `header.ascx.cs` file, which has been changed to create more readable titles. However, the integration package will still work without recompilation.)

5. Restart your Web server.

## Installing on Your Own Custom Site

When you are ready to install the package on your own, customized web site, you should install the same files as described above for the demo site, but you need to be more careful to preserve your site's existing default and header files. The full procedure is given below.

1. Back up your entire web site.
2. Extract the MondoSearch for EPiServer package and locate the following folder:

---

```
<IntegrationPackageRoot>\SamplePack\webroot_460_and_461
```

---

Copy the entire contents of this folder to your EPiServer Web root.

**IMPORTANT:** the following files are included in the package and are replacements for optimizing the standard EPiServer demo site. *Do not overwrite your own versions of these files!*

- <EPiServerSiteRoot>\default.aspx
- <EPiServerSiteRoot>\templates\units\header.ascx
- <EPiServerSiteRoot>\templates\units\header.ascx.cs

Where <EPiServerSiteRoot> is the root content folder for your site.

3. If you are running EPiServer 4.60 then delete the following file (which holds the language cache):

---

<EPiServerSiteRoot>\util\temp\<SiteName>.cache

---

Where <SiteName> is an (often long) name generated by EPiServer. Among other things, the name incorporates the domain name of your site.

(This step enables the language part of the package to be available and preserves the look of the edit/admin GUI in EPiServer 4.60. The problem that makes this step necessary has been fixed in EPiServer 4.61.)

4. Restart your Web server.

## Required Initial Configuration Settings

Though all of the required files should now be in place on your server(s), you still need to make a few configuration settings before the integration will function correctly. Follow each of the procedures below to finish setting up your integration.

### Configuring MondoSearch

**NOTE:** if you are using a hosted MondoSearch installation, then some or all of the steps described in this section might already have been done for you. Please contact your hosting provider to find out which of these are necessary for you. Alternatively, you could work through each of the procedures below and confirm each setting.

If you have installed your Mondosoft applications locally (or if you want to double-check your settings on a hosted installation), do the following:

1. Log into the MondoSearch InSite. (If necessary, contact your hosting provider or system administrator for the required URL, user name and password.)



2. Select **CRAWLER SETUP > CATEGORIES > CatMap Settings** from the InSite navigator bar on the left side of the screen.

Mark the **Allow New Categories to be Created by META-tags** check box and click on **Save**. This will enable new MondoSearch categories to be created directly in EPi-Server without needing to open the InSite.

3. Click on **Done**.
4. Select **CRAWLER SETUP > META-TAGS > Define** from the InSite navigator bar.

#	Name	Type
1	DESCRIPTION	Description
2	DC.Description	Description
3	KEYWORDS	Keywords
4	DC.Subject	Keywords
5	ROBOTS	Robot exclusions
6	LANGUAGE	Language
7	DC.Language	Language
8	CONTENT LANGUAGE	Language
9	TITLE	Title
10	DC.Title	Title
11	CATEGORY	Category
12	MAIN.DOCUMENT	Main document
13	LINK.LIST	Link list
14	KEEP.DOCUMENT	Keep document
15	URL-TO-SHOW	URL to show

Here, we must configure MondoSearch to recognize two custom META-tags that are supplied by EPiServer.

*Do the following:*

- Enter “EPISERVER\_PAGENAME” in the **Name** field at the top of the page.
- Set the **Type** drop-down list to “Title”.
- Click on the **Add** button.

*Do the following:*

- Enter “EPISERVER\_PAGELANG” in the **Name** field at the top of the page.
- Set the **Type** drop-down list to “Language”.
- Click on the **Add** button.

*Do the following:*

- Enter “EPISERVER\_MSALLPAGES” in the **Name** field at the top of the page.
- Set the **Type** drop-down list to “Keywords”.
- Click on the **Add** button.

Click on **Save** to apply your new settings.

Note that MondoSearch also makes use of other META-tags supplied by EPiServer (including “CATEGORY” and “KEYWORDS”) but these are already part of the default configuration.

5. Click on **Done**.
6. Select **CRAWLER SETUP > OPTIONS > InSite Settings** from the InSite navigator bar. Make sure that the **Generate Search Log** check box is marked on this page.

**NOTE:** if some or all of your EPiServer site is protected behind log-in forms, then you must also configure the MondoSearch Crawler to be able to get past these forms if you want to index the content behind them. See “Configuring Crawler Start Points and Log In” on page 31 for details about how to do this.

## Configuring BehaviorTracking

If you are using a hosted MondoSearch and BehaviorTracking service, then this step has already been done for you. However, if you are using a local MondoSearch installation, then you must configure the BehaviorTracking Importer program to locate and import the MondoSearch search log at regular intervals. This requires two steps:

1. On the machine where you are running the import program, use the Windows **Scheduled Tasks** control panel to run the import program at an appropriate interval.
2. Log into the native BehaviorTracking site and go to the **Administration > Site Administration** page. In the **Log-File Paths** field, enter the path to the folder(s) where MondoSearch stores its log files.

See the *BehaviorTracking User’s Guide* for complete details about how to perform each of the above steps.

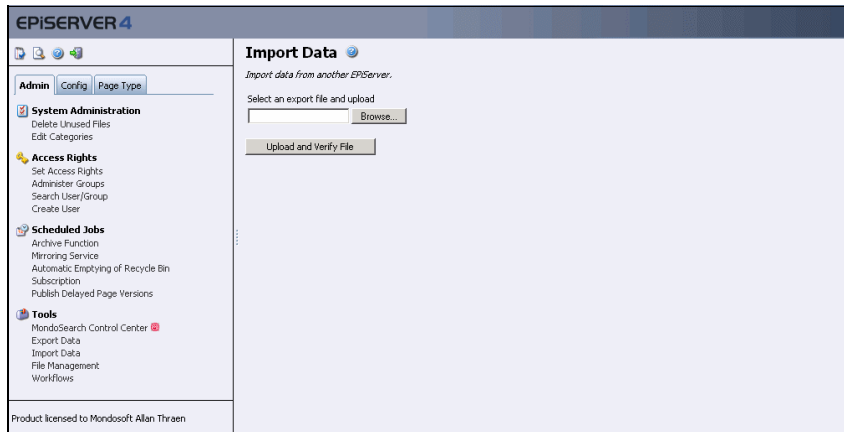
## Configuring InformationManager

The InformationManager Web service requires that you create at least one “local” InformationManager user. We recommend that you use the same login and password as the MondoSearch InSite user.

Again, if you are using a hosted MondoSearch and InformationManager service, then this has already been set up for you. However, if you have a local installation, then please see the *InformationManager User's Guide* (near the end of the “Installing Information-Manager” chapter) for details about how to create and manage users in Information-Manager.

## Configuring EPiServer

1. Open your EPiServer site in a browser. Log in as a user with administration privileges and go into Admin Mode (e.g., by using the right-click context menu).
2. From the **Admin** tab in the left frame, choose **Tools > Import Data**.



Click on the **Browse** button to navigate to and select the following file:

---

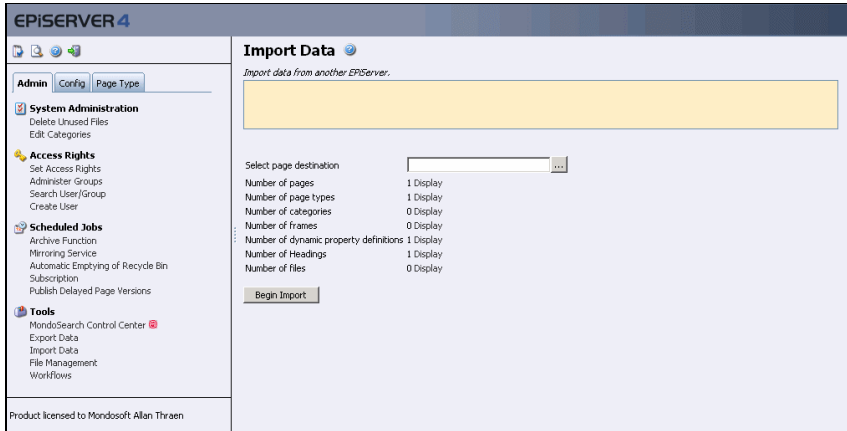
`<IntegrationPackageRoot>\SamplePack\importfiles\  
MondoSearch Integration Package.epi4`

---

Where `<IntegrationPackageRoot>` is the folder to which you unzipped the package.

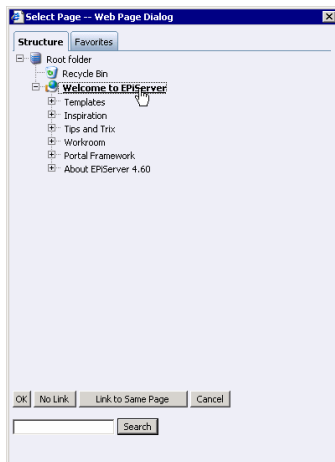
**Note:** Among other things, this import file adds the following to your EPiServer installation:

- New page type: “MondoSearch page”
  - New page: “MondoSearch”
  - New dynamic property: “MondoSearchCategory”
  - New custom property type: “MondoSearchSearchOptions”
3. Click on the **Upload and Verify File** button.



The files are now verified and listed.

4. Click on the ellipsis button next to the **Select page destination** field to open the **Select Page** window.



Select the start page (not the site-wide root) of your EPiServer site and click on **OK** to go back to the **Import Data** page.

5. Click on **Begin Import** to import the files.

- From the **Config** tab in the left frame, choose **Tool Settings > MondoSearch Configuration**. Then go to the **Web Service Settings** tab and enter connection details for each of the various MondoSearch applications.

The screenshot shows the EPiServer 4 administration interface. On the left is a navigation tree with 'Admin', 'Config', and 'Page Type' tabs. Under 'Config', 'System Configuration' is expanded, showing 'System Settings', 'Security', and 'Tool Settings'. 'Tool Settings' is further expanded to show 'MondoSearch Configuration'. The main content area is titled 'MondoSearch Configuration' and has a subtitle 'Use the MondoSearch Configuration settings to set up your integration to MondoSearch™, BehaviorTracking™ and InformationManager™'. It features several tabs: 'General Settings', 'Web Service Settings' (which is active), 'Crawler Settings', 'Background Job', and 'Diagnostic'. The 'Web Service Settings' tab contains several sections for configuring web services. Each section has a 'MondoSearch' URL and a 'connection string' field. The 'MondoSearch' URL field is pre-filled with 'http://episerverdemo.mondosearch.com/SearchService/SearchService'. The 'connection string' field is pre-filled with 'USERNAME="epi" HOSTNAME="episerverdemo.mondosearch.com" CRYPT=MjSw+tk'. There are 'Create New...' buttons next to each connection string field. At the bottom of the page, there is a 'Save' button, a warning message 'Warning! Saving these settings will restart the web application.', and a 'Fill with Demo Settings' button.

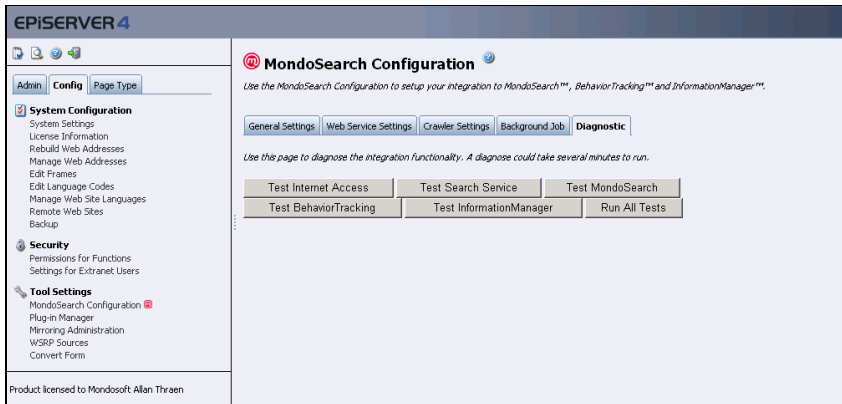
See “Configuring Web Service Communications” on page 49 for complete details about how to enter the information required here.

When you are done making the settings, click on the **Save** button.

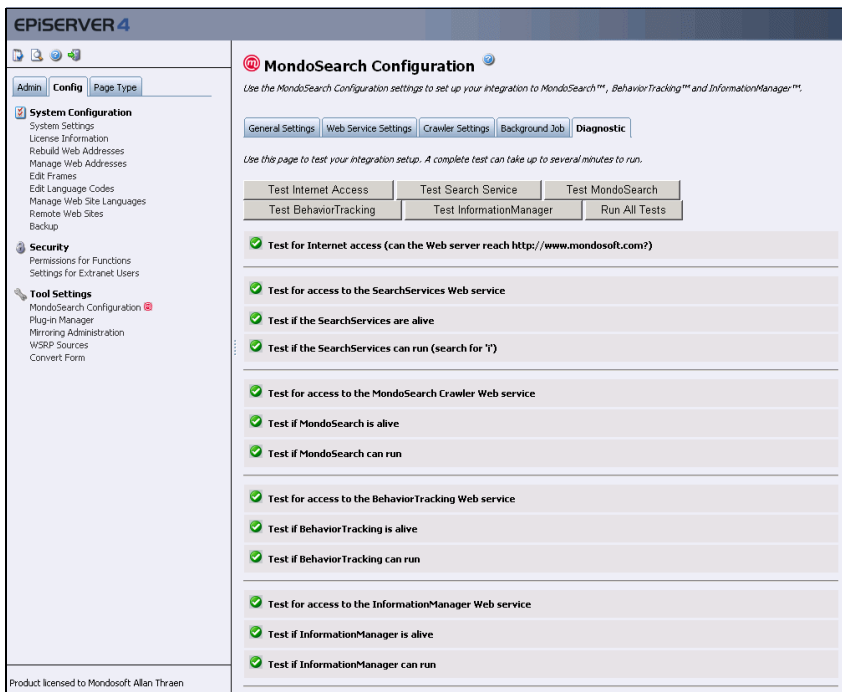
Note that you can fill this page with demo settings by clicking on the **Fill with Demo Settings** button. This will configure your installation to use a demo service placed at Mondosoft. The demo enables you to preview some of the MondoSearch functionality, but not all features and settings will be available.

Note also that if you have not licensed BehaviorTracking and/or InformationManager for your MondoSearch installation, then you can simply leave the URL and connection string settings for these applications blank. The associated functionality will similarly be unavailable in EPiServer.

- From the **Config** tab in the left frame, choose **Tool Settings > MondoSearch Configuration**. Then go to the **Diagnostic** tab.

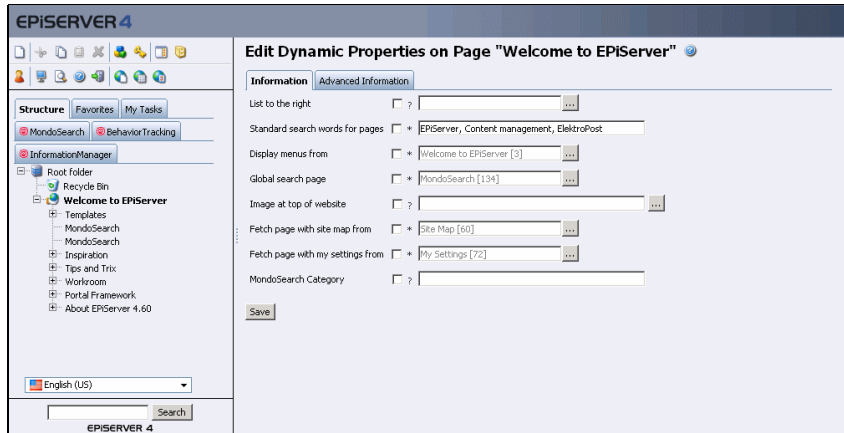


Click on the **Run All Tests** button. In can, in some cases, take a minute or two to run all of the tests. When the results are returned, inspect the diagnostics display to confirm that EPiServer has access to all of the Web services and that all services are alive.



Note that some of the tests will fail for new installations because there is no data in the search or BehaviorTracking database. If one or more of the service-access or “is alive” tests fails, then check your settings for the associated Web service and try again. See also “Using the Diagnostics Tool” on page 56 for more information about this tool.

8. Restart your Web server once again.
9. Return to the EPiServer site in your browser. Log in as a user with editing privileges and go into Edit Mode (e.g., by using the right-click context menu).
10. From the **Structure** tab in the left frame, select the top-level (start) page of your site.
11. Click on the **Edit dynamic properties** button in the edit toolbar (the button looks like three colored cubes stacked into a pyramid).



For the **Global search page** setting, do the following:

- Click on the ellipsis button to open a browse window. Use this window to select the new **MondoSearch** page from your page structure.
- Mark the check box next to this input field. This will ensure that all pages in your site will be updated to use MondoSearch.

Click on **Save** to apply your settings.

12. If you are working with the standard EPiServer demo site, then the code required to place the MondoSearch META-tag emitter is already part of all of the pages of the demo site. However, if you are installing on your own customized site, then you must add this code to the site in a way that will ensure that all pages inherit the META-tag emitter in the <HEAD> section of each page. Please refer to “Placing the META-Tag Emitter on Your Pages” on page 39 for guidelines.

## Configuring Crawler Start Points and Log In

### Crawler Settings on the InSite

If some or all of your EPiServer site is protected behind log-in forms, then you must configure the MondoSearch Crawler to be able to get past these forms if you want to index the content behind them. Usually, you should configure the Crawler to log in with an account that has full access to all of the content areas of your site. When searching, MondoSearch can then be set to filter results to ensure that they only include pages to

which the current searching user has access (see also Chapter 9: “Providing Authorized Result Lists”).

**NOTE:** if you are using a hosted MondoSearch installation, then some or all of the steps described in this section might already have been done for you. Please contact your hosting provider to find out which of these are necessary for you. Alternatively, you could work through each of the procedures below and confirm each setting.

If you have installed your Mondosoft applications locally (or if you want to double-check your settings on a hosted installation), do the following to configure the Crawler with a starting point and to enter user settings for the login system most often used on EPiServer sites:

1. Log into the MondoSearch InSite. (If necessary, contact your hosting provider or system administrator for the required URL, user name and password.)
2. Select **Crawler setup > HOSTS > GrabList** from the InSite navigator bar on the left side of the screen.

The screenshot shows the MondoSearch InSite interface. On the left is a sidebar with a navigation menu. The main content area is titled "GrabList" and contains a table with the following columns: #, URL, Starting Point, and Index, Follow. The first row of the table has the URL "http://ukraine.dev.mondosoft.com/" and the Starting Point checkbox is checked. Below the table are several empty rows for adding more hosts. At the bottom of the interface are navigation buttons: "<<", "<", ">", ">>", "Save", and "Done".

Here you probably already see the host name for your EPiServer site; this is configured when you install MondoSearch. Do the following:

- Click on the **Edit** button next to the listed host name (if no host name is listed, then skip this step and use the blank field at the top of the list instead).
- In the input field, enter the full URL to your site’s login page and include a parameter to redirect to your site’s actual front page. For example:  
`http://www.mysite.com/Templates/login.aspx?ReturnUrl=/Default.aspx`  
(Please see your EPiServer documentation for details about how to find your login page and how to redirect to the front page.)
- Make sure the **Starting Point** check box is marked.
- Set the drop-down list to “Follow”.

Depending on you have structured your site, you may need to add other starting points here. See also the MondoSearch User’s Guide for details about how to work with starting points and fine-tune the Crawler to reach all of the required areas of



your site. Usually, however, the above starting point setting is the only one you will need.

3. Click on **OK** to submit your new starting point settings. Then click on **Save** at the bottom of the page. (Your setting is not saved until you do this!)
4. Click on **Done**.
5. Select **Crawler setup > HOSTS > Content** from the InSite navigator bar on the left side of the screen.

Information	
Company	MondoSoft A/S
Version	5.2
Install date	12/12/2001
Max. hosts	32
Max. pages	50000
User ID	SC
Status	Crawler Resting
License Features	Stat, PDF, MS Office, Flash
Related Links	InformationManager BehaviorTracking

Content	
<b>Grab Link Lists</b>	<input checked="" type="checkbox"/>
Check to grab link lists: <a href="#">[Online Help]</a>	
<b>Follow Cookies</b>	<input checked="" type="checkbox"/>
Check to follow cookies: <a href="#">[Online Help]</a>	
<b>Follow Empty Links</b>	<input checked="" type="checkbox"/>
Check to follow empty links: <a href="#">[Online Help]</a>	
<b>Follow Image Map Links</b>	<input checked="" type="checkbox"/>
Check to follow image map links: <a href="#">[Online Help]</a>	
<b>Follow Image Links</b>	<input checked="" type="checkbox"/>
Check to follow image links: <a href="#">[Online Help]</a>	
<b>Follow GET Form Links</b>	<input type="checkbox"/>
Check to follow GET form links: <a href="#">[Online Help]</a>	
<b>Follow POST Form Links</b>	<input checked="" type="checkbox"/>
Check to follow POST form links: <a href="#">[Online Help]</a>	
<b>Split Pages at Bookmarks</b>	<input type="checkbox"/>
Check to divide pages into several subpages: <a href="#">[Online Help]</a>	

Make sure that the following two check boxes are marked:

- **Follow Cookies**  
This setting makes sure that the Crawler will store and return cookie values as it logs in and crawls the EPiServer site. This is required for, among other things, maintaining an authenticated user session.
- **Follow POST Form links**  
This setting enables the Crawler to submit the login form and continue on to crawl the site.

6. At the bottom of the **Content** page, click on the **Edit** button next to the **Form Field** heading. This brings you to a form for defining values that will be submitted to the EPiServer login form.

#	Form Field Name	Form Field Value
	<input type="text"/>	<input type="text"/>

[Online Help]

7. Do the following:
  - In the **Form Field Name** field, enter “username”.  
(This is the name of the user name field in most EPiServer installations; if your site includes a customized form where this value is changed, then you may need to enter a different value here.)
  - In the **Form Field Value** field, enter the user name for the account that should be used by the Crawler as it indexes your site. As mentioned above, you should use an account that is able to access all of the pages that you want to include in the search database.

Click on the **Add** button.

8. Do the following:
  - In the **Form Field Name** field, enter “password”.  
(Again, this is the usual name for the password input field; it is possible (though unlikely) that you should use another name for your site.)
  - In the **Form Field Value** field, enter the password for the account you named in the previous step.

Click on the **Add** button.

9. Click on **OK** to return to the previous page. (**Important:** your settings are still not saved yet!)
10. Click **Save** on the Content page to save your new cookie, form and value settings.

## Modify the Robot Commands on Your Login Page

The default login page supplied with EPiServer contains the following META-tag:

---

```
<meta name="robots" content="noindex,nofollow">
```

---

This tag instructs all Web crawlers (including MondoSearch) to neither index the page nor follow its links. However, we want the MondoSearch Crawler to submit the form and follow the link, so you must change this setting if your login form has it. Open the login page you are using for your site in a text editor and look for this META-tag in the <HEAD> section of the page. If it is present, change it to the following:

---

```
<meta name="robots" content="noindex,follow">
```

---

## Testing the Installation

Once you have made the basic installation and configuration settings described in this chapter up to this point, then you are ready to make your first test of the system. If you would like to do so now, then start the Crawler manually and, once crawling is finished, try to search from your site (see Chapter 6: “Manual Crawler Control and Monitoring”). Once your solution is fully up and running, you will probably schedule the Crawler to run automatically at an appropriate interval (see also “Scheduling the Crawler” on page 53).

You will need to wait until some usage data has been collected and the BehaviorTracking Importer has run before you will be able to access BehaviorTracking data from within EPiServer. This might take a day or two.

## Blocking Unneeded CGI Access to MondoSearch

By default, MondoSearch generates CGI-based search/result pages on its server after each crawl. Because you are using EPiServer, it is likely that you will not need these pages at all. To enhance the security of your solution, consider configuring the IIS installation on your MondoSearch server so that it will not execute these pages (or asking your hosting provider to do this for you). One way to do this is to configure IIS so that it never runs the MondoSearch CGI Finder program (*MsmFind.exe*).

Among other things, the CGI-based pages do not apply authorization to result lists, so if you are using the authorization feature then you should take steps to disable the CGI pages.





# CHAPTER 3

## The META-Tag Emitter



Many of the advanced features of MondoSearch are made possible by the search engine's ability to index and process special types of META-tags, which provide extra information about each page. Many of these META-tags are added to the page automatically by a META-tag Emitter control, which is provided with the integration package. Among other things, this control helps implement the categorization and result-list authorization features.

This chapter describes the MEAT-tag emitter control and explains how to use it.

## Purpose of the META-tag Emitter

Several features of the integration, including categorization and Result-set filtering, are made possible by a META-tag emitter added to EPiServer when you install the MondoSearch integration package. This control generates several META-tags, which must be placed in the <HEAD> section of each rendered HTML page.

For example, to enable the result-list authorization feature, these tags provide a list of user groups that have access that page. As a result, this information is available in the search database for each page, so it is very easy for MondoSearch to filter results so that they include only authorized pages in the result list.

## META-Tags Added by the Emitter

The META-tag emitter generates the following tags (identified below with the name property established for each META-tag by the emitter):

- **EPISERVER\_MSALLPAGES**  
This tag always has the same value (“MSALLPAGES”). Because these value then becomes part of every page, it becomes possible to find all of the pages in the database by searching for this value. For this to work, you must configure MondoSearch to recognize this tag as a keywords tag (see also “Configuring MondoSearch” on page 24).
- **EPISERVER\_PAGENAME**  
Provides the title of each page, as shown in search results. For this to work, you must configure MondoSearch to recognize this tag as a title tag (see also “Configuring MondoSearch” on page 24).
- **EPISERVER\_PAGEID**  
Provides the unique ID number for the page, as it is known internally in EPiServer. Among other things, this enables the MondoSearch Editor Search feature to open found pages in edit view (see also Chapter 10: “Editor Search”).
- **EPISERVER\_PAGELANG**  
Identifies the language of the page. MondoSearch displays this in search results and also uses it when applying advanced language processing such as phonetics and stemming. For this to work, you must configure MondoSearch to recognize this tag as a language tag (see also “Configuring MondoSearch” on page 24).
- **EPISERVER\_BREADCRUMB**  
Identifies the full path through the site structure that leads to the current page (as shown in EPiServer Edit Mode). This is not used in the default solution, but you could customize your result page to show this value for each found document.
- **EPISERVER\_READACCESS**  
This tag reports a series of user-group IDs that have access to the page. If authentication is enabled, then only those pages that include the searching user’s user group here will be included in search results.
- **EPISERVER\_CATEGORIES**  
This tag reports the EPiServer category (or categories) to which the page belongs. Usually, these categories are not used by MondoSearch (which provides an alternative

categorization system). However, the first of the categories listed here might sometimes be copied to the “CATEGORY” tag if no MondoSearch category is available (see the next point, below). Note that you should *not* configure MondoSearch to interpret this tag as a category tag; the “CATEGORY” tag is used for this purpose (again, see the next point, below). You might also use the value of this tag to create custom filters (see the *MondoSearch Developer’s Guide* for details about custom filters).

- **CATEGORY**

This is the category that will actually be used in MondoSearch. It is calculated as follows: if the page has a property or dynamic property called “MondoSearchCategory” then that value is taken (see also “Making Category Assignments” on page 73); otherwise, first category identified for EPISERVER\_CATEGORIES is used (if present); if neither type of category is available, then this tag will not be added to the page.

## Placing the META-Tag Emitter on Your Pages

When you install the integration package on the default EPiServer demo site, the files copied to the site automatically include the code required to place the MondoSearch META-tag emitter on each page. However, when you want to integrate with your own, custom site, you must incorporate this code in a way that will ensure that it is placed on every page. Usually you will do this by establishing a file that generates content for the <HEAD> section of each page and then design all other page types to inherit this file.

The code required to place the MondoSearch META-tag emitter is:

---

```
<%@ Register TagPrefix="MondoSoft"
Namespace="Mondosoft.EPiServerIntegration.WebControls"
Assembly="MondoSoft.EPiServerIntegration" %>
.
.
.
<MondoSoft:MetaData id="MetaData" runat="server" />
```

---

A detailed description of how to work with this type of coding in .NET and EPiServer is beyond the scope of this manual. Please refer to your EPiServer and/or .NET documentation for additional assistance. To see how this is done for the default EPiServer demo site, install the integration package on the demo site and examine the following file:

---

```
<EPiServerSiteRoot>\templates\units\header.ascx
```

---

## Controlling Access to the Metadata

Usually, you will configure the META-tag emitter control so that it only delivers its extra META-tags to the MondoSearch crawler and not to other search engines or normal users. This will help ensure that sensitive information, such as your user group names, does not get delivered to unauthorized parties. It will also deliver a small performance increase.

However, you might also decide to deliver the META-tags to all clients. See “Setting the Crawler Agent ID” on page 55 for details about how to configure this option.





# CHAPTER 4

## Optimizing Your Site for Search



For most sites, MondoSearch will automatically provide excellent results right out of the box. However, there are a few ways that you can optimize your site to help MondoSearch work even better. Though the most important steps for doing this are included in Chapter 2: “Installing MondoSearch for EPiServer”, this chapter gives a few more ideas for how you can customize your site’s pages to help make sure your users find what they are looking for quickly and easily.

## Optimizing Page Titles and Descriptions

Ensuring that your site provides accurate page titles and descriptions is probably the most important thing that you can do to provide useful result lists from your search engine.

In the default integration setup, MondoSearch uses the page titles as they appear in EPiServer (these are fetched by the META-tag emitter; see also “The META-Tag Emitter” on page 37). In EPiServer, inspect your titles to decide if they are descriptive enough and update them as necessary. Alternatively, you could customize the search/result page to use the MondoSearch feature for generating titles based on page content. To enable this feature, configure your search/result page with **Additional MQL** settings, such as:

---

```
OPTIONS TitleMode = 'TextTitle ', UniqueTitle='true'
```

---

Similarly, MondoSearch displays a description for each document listed on the single-category result page. However, these are not fetched automatically in the default setup (though when you install on the EPiServer demo site, a sample solution is provided; see “Installing on the Standard EPiServer Demo Site” on page 22). To supply descriptions for your own custom site, you could either modify your site to include “DESCRIPTION” META-tags on each page or, as above, configure your search/result page with **Additional MQL** settings that will use auto-generated descriptions. For example:

---

```
OPTIONS DescriptionMode = 'TextDesc ', UniqueDescription='true'
```

---

For details about how to add MQL to your search/result page, see “Configuring Search Options” on page 66. For complete details about the available MQL syntax (including a few other options for generating titles and descriptions), see the *MondoSearch Developer's Guide*.

Note also that another way to implement the above features is to modify the search/result .NET page directly using (for example) Microsoft Visual Studio. The .NET page supplied for EPiServer is very similar to the sample template included with MondoSearch. For details about how to customize this page, please see the *MondoSearch Developer's Guide*.

## Improving Page Rankings

Usually, the advanced, context-sensitive ranking system used by MondoSearch will give excellent results for all sites. However, you could also fine-tune the ranking response for your site by adding META-tags to specific pages that will cause them to rank higher, rank lower and/or appear in searches where they might not otherwise be included. MondoSearch supports the following tag types for these purposes:

- *Rank-word tags*: cause the page to rank much higher when a user searches for one of the words defined by this tag. (Note, however, that the word must also appear in standard text on the page, or it will be ignored.)

- *Rank-document tags*: enable you to set the page to rank either higher or lower than it would otherwise in all searches where it is found.
- *Keyword tags*: enable you to add words that do not otherwise appear on the page, causing the page to be listed in searches for those words. Keywords will also have a slight affect on ranking.

To use these, you must customize your page types to enable authors to enter values for these META-tags. Please see the *MondoSearch User's Guide* for complete details about naming and configuring these tags.

Often, data from BehaviorTracking will help you find out when and where you should apply these tags. See also Chapter 11: "Working with BehaviorTracking in EPiServer".

## Marking Content for Exclusion from the Search Database

If your site includes pages (or parts of pages) that should not be in the search database, then you can prevent these from being indexed by embedding crawler commands in your documents. For example:

- If your site includes printable layouts for content normally shown on-screen, you can mark these pages to be ignored by the Crawler using robot-control META-tags.
- If your pages include sections that should not be indexed, then enclose these in `<NOINDEX>` tags. Use these, for example, to keep text from navigator links from being indexed as content on the page even while allowing the links to be followed.
- If your page includes links that should not be followed, then enclose these in `<NOFOLLOW>` tags. This is another way to keep printable versions out of the search database.

These techniques are described in more detail in the sub-sections below.

Note that MondoSearch usually identifies pages that have duplicate content, even when the URL and layout are different. This will help ensure that result lists do not show both the printable and online versions of the same information. However, the techniques outlined above (and explained below) will give you better control.

### Excluding Entire Documents

Use the robot-control META-tag to control whether a page should be grabbed and/or whether its links should be followed. Though you can configure MondoSearch to recognize any named META-tag for this purpose, most other search engines recognize this type of tag by the NAME "robots" (and MondoSearch recognizes this name by default). The CONTENT value of the tag is a comma-separated combination of one or more of the following values:

- **index** | **noindex**: specifies whether or not the page should be grabbed (indexed)
- **follow** | **nofollow**: specifies whether or not the page's links should be followed
- **all**: specifies that the page should be grabbed and all links should be followed

- **none:** specifies that the page should not be grabbed and the links should not be followed.

**Example:**

The following META element tells search engines and other robots not to grab (index) the page but to follow links on it:

---

```
<META NAME="ROBOTS" CONTENT="noindex, follow">
```

---

**NOTE:** another way to mark pages and/or folders of your site for exclusion from the database is to use the MondoSearch InSite GrabMap, which shows a representation of your site as it appears to the MondoSearch Crawler and enables you to make settings and commands. See the *MondoSearch User's Guide* for complete details about this and other features that enable you to fine-tune the way the Crawler accesses your site.

## Excluding Parts of Documents

The MondoSearch Crawler recognizes two types of special HTML tag pairs invented by Mondosoft. Page authors can use these tag pairs to exclude sections of text and/or links from a page otherwise indexed by the Crawler. This technique will have no effect on other search engines.

Text placed between `<NOINDEX>` tags will not be saved in the database, but links contained there may be followed. Links placed between `<NOFOLLOW>` tags will not be followed by the crawler, but surrounding text may be indexed. Nest these two tags to mark sections that should be neither followed nor indexed.

**Syntax:**

---

```
<NOINDEX> ungrabbed text </NOINDEX>  
<NOFOLLOW> ...unfollowed links... </NOFOLLOW>
```

---

Alternatively, you can use HTML comments to enclose your no-index and no-follow tags. This approach conforms more exactly to HTML standards.

**Alternative Syntax:**

---

```
<!--NOINDEX--> ungrabbed text <!--NOINDEX-->  
<!--NOFOLLOW--> ...unfollowed links... <!--NOFOLLOW-->
```

---



# CHAPTER 5

## MondoSearch Administration



The MondoSearch integration package installs an administration interface that enables you to administer and troubleshoot several of the technical aspects of the integration, including Web-service connection details and the Crawler schedule. Usually, you will use these just once as you set up and test the integration, though they are available to administrators for use at any time.

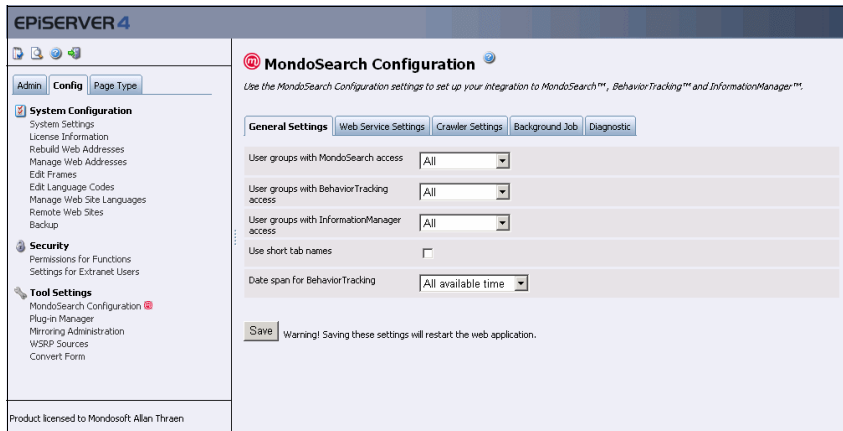
## Accessing the MondoSearch Administration Settings

To access the MondoSearch administration settings:

1. Log in to EPiServer as a user that has administrator privileges.
2. Enter Admin Mode (e.g., by using the right-click context menu or clicking on the **Admin** button while in Edit Mode).
3. Click on the **Config** tab in the left frame.
4. Click on the **MondoSearch Configuration** link under the **Tool Settings** heading in the left frame.
5. Use the various tabs of the **MondoSearch Configuration** page in the right frame to navigate to the various administration settings. All of the available settings are described in the remaining sections of this chapter.

### “General” Settings

When you first come to the **MondoSearch Configuration** page, you will see the **General Settings**. These included several miscellaneous options for controlling the appearance and availability of MondoSearch features within EPiServer. You can also set the time span of data included in the BehaviorTracking displays.



### Controlling Edit-Mode Features' Appearance and Availability

By default, the MondoSearch integration package displays BehaviorTracking information, InformationManager controls and page-search features to all users that have access to Edit Mode in EPiServer. However, if you wish to hide some or all of these features, you can configure the solution to do so.

To set these options:

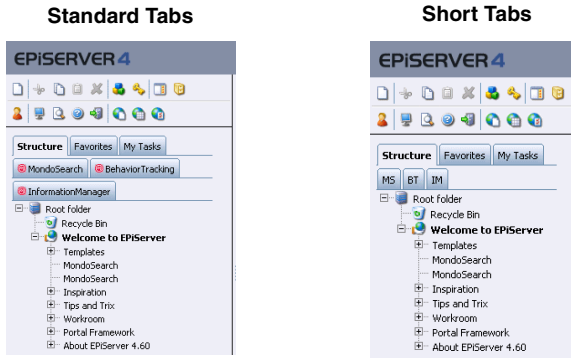
1. Go to the **MondoSearch Configuration** page and select the **General Settings** tab (see also "Accessing the MondoSearch Administration Settings" on page 46).

The screenshot shows the EPiServer 4 administration interface. On the left is a navigation menu with categories: System Configuration (containing System Settings, License Information, Rebuild Web Addresses, Manage Web Addresses, Edit Frames, Edit Language Codes, Manage Web Site Languages, Remote Web Sites, Backup), Security (Permissions for Functions, Settings for Extranet Users), and Tool Settings (MondoSearch Configuration, Plug-in Manager, Mirroring Administration, WSRP Sources, Convert Form). The main content area is titled 'MondoSearch Configuration' and includes a subtitle: 'Use the MondoSearch Configuration settings to set up your integration to MondoSearch™, BehaviorTracking™ and InformationManager™'. Below this are tabs for 'General Settings', 'Web Service Settings', 'Crawler Settings', 'Background Job', and 'Diagnostic'. The 'General Settings' tab is active, showing three dropdown menus: 'User groups with MondoSearch access' (set to 'All'), 'User groups with BehaviorTracking access' (set to 'All'), and 'User groups with InformationManager access' (set to 'All'). There is also a checkbox for 'Use short tab names' (unchecked) and a dropdown for 'Date span for BehaviorTracking' (set to 'All available time'). At the bottom of the settings area is a 'Save' button and a warning message: 'Warning! Saving these settings will restart the web application.' The footer of the page states 'Product licensed to Mondosoft Allan Thraen'.

2. Set the following drop-down lists as needed:
  - **User groups with MondoSearch access:** Select the user group that should be able to access the MondoSearch search form while in Edit Mode (MondoSearch will always be available on the main Web site). Normally, this search form makes it easier for editors to find the pages that they want to work with, but you might decide to hide this form in some situations (such as if your site contains fairly few pages). See Chapter 10: "Editor Search" for more information about this feature.
  - **User groups with BehaviorTracking access:** Select the user group that should be able to see BehaviorTracking information in Edit Mode. Regardless of this setting, the data will still be gathered and the front-page display of most popular search selections will still be available. See Chapter 11: "Working with BehaviorTracking in EPiServer" for details about the features available.
  - **User groups with InformationManager access:** Select the user group that should be able to see InformationManager controls in Edit Mode. See Chapter 12: "InformationManager Features in EPiServer" for details about the features available.

Each of the above drop-downs includes the name of each user group currently defined in EPiServer and also includes entries for "All" and "None", which make the feature available to all users or no users, respectively.

- Each of the features added to Edit Mode has its own tab in the left frame of the Edit-Mode interface. By default, these are labelled with the full name of each feature (MondoSearch, InformationManager and BehaviorTracking). If, however, you prefer to conserve space in the interface (at the expense of somewhat more vague tab labels) mark the **Use short tab names** check box.



- Click on **Save** to apply your new settings.

## Setting the BehaviorTracking Date Span

The various BehaviorTracking views shown in the Edit-Mode interface display statistics about the way users have been searching on your site. Though the BehaviorTracking database might include data spanning many months, or even years, the most relevant statistics will often be those based on data collected in only the last few weeks or months. To select the range of dates for which you would like to view statistics:

- Go to the **MondoSearch Configuration** page and select the **General Settings** tab (see also "Accessing the MondoSearch Administration Settings" on page 46).
- Use the **Date span for BehaviorTracking** drop-down list to select a date range. In addition to the self-explanatory selections (such as "Previous 90 days") is an entry marked "default"; this entry uses the primary date span currently configured for the user account used by EPiServer in the BehaviorTracking installation. See the *BehaviorTracking User's Guide* for details about how to establish this date span.
- Click on **Save** to apply your new setting.



## Configuring Web Service Communications

All of the interactions between EPiServer and the MondoSearch Suite applications are made via the following Mondosoft Web services:

- **MondoSearch SearchService:** processes queries to the search database and returns search results.
- **MondoSearch Maintenance Service:** accepts commands to start, stop or kill the Crawler and returns Crawler status.
- **BehaviorTracking Web service:** processes queries to the BehaviorTracking database.
- **InformationManager Web service:** assigns and returns SearchHeaders, SearchNames and synonyms.

For EPiServer to access each of these services, it must be configured with the following:

- The URL at which the service is provided.
- A connection string, which includes user name, encrypted password and/or license code (depending on the service).

To make these settings, go to the **MondoSearch Configuration** page and select the **Web Service Settings** tab (see also “Accessing the MondoSearch Administration Settings” on page 46). Refer to the sub-sections below for details about how to find and enter the required values for each service.

The screenshot shows the EPiServer 4 administration interface. The left sidebar contains navigation links: Admin, Config (selected), and Page Type. Under Config, there are sections for System Configuration, Security, and Tool Settings. The main content area is titled 'MondoSearch Configuration' and includes a sub-header: 'Use the MondoSearch Configuration settings to set up your integration to MondoSearch™, BehaviorTracking™ and InformationManager™.' The 'Web Service Settings' tab is active, showing configuration fields for three services: MondoSearch, BehaviorTracking, and InformationManager. Each service has fields for a connection string (with a 'Create New...' button) and a web service URL. At the bottom, there are 'Save', 'Warning! Saving these settings will restart the web application.', and 'Fill with Demo Settings' buttons.

Service	Connection String	Web Service URL
MondoSearch	7d4248463149916c73902d4a66b3513f	http://episerverdemo.mondosearch.com/SearchService/SearchService
BehaviorTracking	USERNAME='epi' HOSTNAME='episerverdemo.mondosearch.com' CRYPT=MjSw+tk	http://episerverdemo.mondosearch.com/SearchService/MSService.as
InformationManager	USERNAME='epi' HOSTNAME='episerverdemo.mondosearch.com' CRYPT=MjSw+tk	http://episerverdemo.mondosearch.com/SearchService/IMService.as

## Web Service URLs

The URLs for each service are determined by the machine on which they are installed. If you are using a hosted version of MondoSearch, then these will be supplied to you by your hosting provider. If you have installed your MondoSearch applications at your own site, then contact your local site administrator if you do not already know your local service URLs.

Once you have found the URLs for each of the four services, simply enter them in the appropriate fields on the **General Settings** tab.

## MondoSearch License Code

Access to the SearchService requires a license code that matches the code of your MondoSearch installation. Because you should keep your MondoSearch license code a secret, you should use an encrypted version of it. Unencrypted codes will also be accepted, but we do not recommend that you use them.

If you are using a hosted version of MondoSearch, then an encrypted version of your MondoSearch license will be given to you by your hosting provider. If you have a local MondoSearch license, then you can use the small command-line encoding utility included with MondoSearch to generate an encrypted version of your license. Using the default installation options, you can find the utility at:

---

C:\MondoSearch\MsmEncryptLicense.exe

---

To generate an encrypted license code, open a command-prompt window and navigate to the folder where the license-encryption program is installed. Then execute the encryption program, adding your un-encrypted license code as an input parameter. For example:

---

MsmEncryptLicense.exe *MyUnecrypedLicenseCode*

---

The program then outputs an MD5 encryption of your code.

Once you have the encrypted version of your MondoSearch license code, simply enter it into the **License for MondoSearch** field on the **General Settings** tab of the **MondoSearch Configuration** page.

The encrypted version of your license code can only be used with the SearchService. You cannot use it directly with MondoSearch (such as when registering the product).

## Connection Strings for the Integration Services

While the SearchService authenticates access using only your MondoSearch license code, the other three services require a connection string that encodes additional information (such as user name, password, etc.). The form of the connection strings varies according to service, as outlined below in Table 1.

Web Service	Required Connection Parameters	Encrypted Connection String Example
MondoSearch SearchService	License	7d4248463149916c73902d4a86b3513f
MondoSearch Maintenance Service	License UserName Password HostName	USERNAME='epi' HOSTNAME='episerver-demo.mondosearch.com' CRYPT='MjSw+tkOAL-hXG/DvpKDCP4/8tR80tArrQK69D34rLbdv/y00mhdzq4GLIUPRLL5I7hwWkntZau7jWCYWvirxHdZgmKPyUz2bGoXLD7ObJSbUmlb7C4G0clVFwURCTNUj6Aa7/E6Q/al='
InformationManager Service	License UserName Password HostName	USERNAME='epi' HOSTNAME='episerver-demo.mondosearch.com' CRYPT='MjSw+tkOAL-hXG/DvpKDCP4/8tR80tArrQK69D34rLbdv/y00mhdzq4GLIUPRLL5I7hwWkntZau7jWCYWvirxHdZgmKPyUz2bGoXLD7ObJSbUmlb7C4G0clVFwURCTNUj6Aa7/E6Q/al='
BehaviorTracking Service	UserName Password	USERNAME='episerverdemo' CRYPT='AgV0GTq65qCYOW9tno7ZtvJBcveU78n2XuQQGPVdtGumjJe49AhvdHlnGeN36YNJfnrMkfvZGZQ='

*Table 1 Examples of encrypted connection strings for each service*

Because the connection strings will very typically be sent over the Internet as plain text, they are encrypted to keep sensitive information, such as passwords and license codes, from being intercepted. Note, however, that the user name and, when present, host name are still sent as plain text. This is required to enable the string to be decrypted when it is received by the providers. The **General Settings** tab provides tools to generate encrypted connection strings for each service.

To generate your connection strings:

1. Collect the required information for each service, as outlined above in Table 1. Note that the various user names and passwords are those configured locally for each MondoSearch application.
  - If you are using a hosted version of MondoSearch, then these will simply be given to you by your hosting provider.

- If you have MondoSearch installed locally, then the user name and/or passwords must be configured for each application after installation. If you were not responsible for the MondoSearch installation, then please contact your system administrator for help.
  - In each case, the license code required is the one for the product associated with each service (BehaviorTracking, MondoSearch or InformationManager).
2. Go to the **MondoSearch Configuration** page and select the **Web Service Settings** tab (see also “Accessing the MondoSearch Administration Settings” on page 46).
  3. For each relevant service, click on the **Create new** button located under that service’s **Connection String** field. This expands the page to include a short form for entering the required information. (Note that there is no **Create new** button for the MondoSearch SearchService; here, you must use the MondoSearch encryption utility to generate an encrypted license code; see “MondoSearch License Code” on page 50 for detail.)

The screenshot shows the EPiServer 4 administration interface. The left sidebar contains a tree view with categories: Admin, Config, and Page Type. Under Config, there are sections for System Configuration (including System Settings, License Information, Rebuild Web Addresses, Manage Web Addresses, Edit Frames, Edit Language Codes, Manage Web Site Languages, Remote Web Sites, and Backup), Security (Permissions for Functions, Settings for Extranet Users), and Tool Settings (MondoSearch Configuration, Plug-in Manager, Mirroring Administration, WSRP Sources, and Convert Form). The main content area is titled 'MondoSearch Configuration' and includes a sub-header: 'Use the MondoSearch Configuration settings to set up your integration to MondoSearch™, BehaviorTracking™ and InformationManager™.' Below this are tabs for General Settings, Web Service Settings (which is selected), Crawler Settings, Background Job, and Diagnostic. The Web Service Settings tab contains several sections: 1. 'MondoSearch license code' with a text field containing '7d4248463149916c73902d4a86b3513f'. 2. 'MondoSearch SearchService URL' with a text field containing 'http://episerverdemo.mondosearch.com/SearchService/SearchService'. 3. A section for 'MondoSearch Crawler Web service connection string' with a note: 'Enter the requested data and click on "Generate" to create the connection string. Don't forget to click on "Save" on the bottom of the page to save the new connection string.' This section includes input fields for User, Password, Host name, and License code, followed by 'Generate Connection String' and 'Cancel' buttons. 4. 'MondoSearch Crawler Web service URL' with a text field containing 'http://episerverdemo.mondosearch.com/SearchService/MSService.as'. 5. 'BehaviorTracking Web service connection string' with a text field containing 'USERNAME="episerverdemo" CRYPT="Agv0GTq65qCYOW9tno7ZhuBcveU78n2XuQZ"' and a 'Create New...' button. 6. 'BehaviorTracking Web service URL' with a text field containing 'http://episerverdemo.behaviortracking.com/WebService/BTService.as'. At the bottom left, it says 'Product licensed to Mondosoft Allan Thraen'.

4. Enter the information requested and click on the **Generate Connection String** button to generate the string.
5. Continue working until you have entered the information required for each service.
6. Go to the **Diagnostics** tab, which enables you to test the connection to each service. If problems are indicated, then go back and re-check the information you have entered for the relevant service(s). See “Using the Diagnostics Tool” on page 56 for details about how to work with the **Diagnostics** tab.

## Demo Settings

It is possible to load demo settings for all of the Web service URLs and connection strings. This will configure your installation to use a demo service placed at Mondosoft. The demo enables you to preview some of the MondoSearch functionality, but not all features and settings will be available. To load the demo settings, click on the **Fill with Demo Settings** button.

## Scheduling the Crawler

When users search your site using MondoSearch, their queries are processed against the MondoSearch search database. This database is not the same as the EPiServer database that holds the content for your site. Unlike the EPiServer database, which is optimized for authoring and serving content, the MondoSearch database is optimized for searching the site as users see it.

The MondoSearch database is created and maintained by the MondoSearch Crawler program, which visits your site like a user does, by requesting and following links. The contents of each page are then indexed and stored in the search database. The search database, therefore, is only as recent as the last Crawler run.

This section explains how to establish a schedule by which to run the Crawler, thereby keeping your search database up to date with your site. Note that you can also run the Crawler manually at any time; see Chapter 6: “Manual Crawler Control and Monitoring” for details about this function.

Crawler scheduling requires two steps:

- Set up a schedule by which to run the crawler
- Set up a background job, which runs every 20 minutes (to check, among other things, whether the Crawler is scheduled to run) and reacts as needed.

## Setting the Crawler Schedule and E-Mail Alerts

To establish the crawler schedule and associated settings:

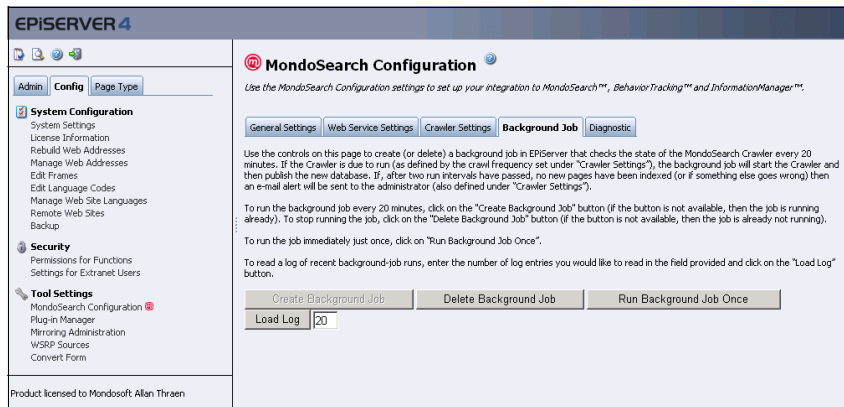
1. Go to the **MondoSearch Configuration** page and select the **Crawler Settings** tab (see also “Accessing the MondoSearch Administration Settings” on page 46).

2. Make the following settings:
  - **Crawl frequency in hours:** enter the number of hours that should pass between each start-Crawler command. A low value here will cause the Crawler to run frequently; a high value will run it less frequently. The Crawler usually generates a large amount of traffic on your network and Web server, so you should select a value that balances your needs for an updated search database with the need to limit Crawler traffic on your network.
  - **Do not start the Crawler within this time frame:** If you wish to prevent the Crawler from starting during times of high network traffic (e.g., during business hours), then specify a do-not-crawl period using these two fields (e.g., 08:00 - 18:00). Note that once the Crawler has started it will run to completion, even if the cut-off time defined here has passed.
  - **Crawler agent for metadata:** You should almost always set this to “MondoSearch”. This setting ensures that the metadata generated for MondoSearch is delivered to the MondoSearch Crawler and to no other crawlers or clients. Clear this field to deliver expanded metadata to all clients. See also “Setting the Crawler Agent ID” on page 55 for more information about this feature.
  - **Send alerts to this e-mail address:** enter the e-mail address of the person assigned to handle issues with EPiServer and MondoSearch. If an error or other unexpected event occurs with the Crawler or MondoSearch installation, an e-mail message will be sent here.
  - **Send alerts from this e-mail address:** this is the address that will appear in the “from” field of the e-mail sent to the address above. This is for the recipient’s reference only; for example, the recipient could configure his or her mail client to filter or flag messages based on this value.
3. Click on **Save** to apply your settings.

## Triggering the Crawl Schedule with the Background Job

Though your Crawler settings establish the crawl frequency and times during which the Crawler should not run, these settings are passive. To make activate them, you must run a background job, which is launched every 20 minutes by EPiServer. At each run, the background job checks a few things related to MondoSearch, including whether the Crawler is scheduled to run, and then acts as required.

To set up the background job, go to the **MondoSearch Configuration** page and select the **Background Job** tab (see also “Accessing the MondoSearch Administration Settings” on page 46).



Work with the controls provided here as follows:

- To run the background job every 20 minutes, click on the **Create Background Job** button (if the button is not available, then the job is running already).
- To stop running the job, click on the **Delete Background Job** button (if the button is not available, then the job is already not running).
- To run the job immediately just once, click on **Run Background Job Once**.
- To read a log of recent background-job runs, enter the number of log entries you would like to read in the field provided and click on the **Load Log** button.

## Setting the Crawler Agent ID

The MondoSearch integration package includes a control that generates several types of META-tags, which provide extra page information that is indexed by the MondoSearch Crawler (see also Chapter 3: “The META-Tag Emitter”). Usually, you should set up your site so that this control is placed into the templates of all EPiServer pages, thereby ensuring that all pages include this enhanced set of META-tags.

Because the META-tags generated by the MondoSearch control are only used by MondoSearch, and also because they can in some cases contain sensitive information (such as user group names), you will usually configure the control only to deliver its META-tags to the MondoSearch Crawler and not to other crawlers or clients. This is possible because most

types of clients identify themselves to the Web server, thereby enabling the server to deliver content most appropriate for each supported client. In the case of MondoSearch, the Crawler reports itself as “MondoSearch-MS/5.2 (<http://mondosoft>)”.

To configure the agent to which the MondoSearch META-tag control should deliver its tags, do the following:

1. Go to the **MondoSearch Configuration** page and select the **Crawler Settings** tab (see also “Accessing the MondoSearch Administration Settings” on page 46).

The screenshot shows the EPiServer 4 administration interface. The left sidebar contains a tree view with categories: Admin, Config, and Page Type. Under Config, there are sections for System Configuration (with sub-items like System Settings, License Information, Rebuild Web Addresses, etc.), Security, and Tool Settings. The main content area is titled 'MondoSearch Configuration' and includes a subtitle: 'Use the MondoSearch Configuration settings to set up your integration to MondoSearch™, BehaviorTracking™ and InformationManager™'. There are five tabs: General Settings, Web Service Settings, Crawler Settings (which is selected), Background Job, and Diagnostic. The Crawler Settings tab contains several fields: 'Crawl frequency in hours' (set to 1), 'Do not start crawl at this time of day (hh:mm - hh:mm)' (empty), 'Crawler agent for Metadata' (empty), 'Send alerts to this e-mail address' (empty), and 'Send alerts from this e-mail address' (empty). At the bottom, there is a 'Save' button and a warning message: 'Warning! Saving these settings will restart the web application.' The footer of the interface states 'Product licensed to Mondosoft Allan Thraen'.

2. In the **Crawler agent for metadata** field, enter the client identifier to which you would like to supply MondoSearch metadata. This will nearly always be the MondoSearch Crawler, which can be specified using either of the following:
  - MondoSearch
  - MondoSearch-MS/5.2 (<http://mondosoft>)

Either value will work the same.

If you prefer to deliver the metadata to all clients and crawlers, then clear this field.

3. Click on **Save** to apply your settings.

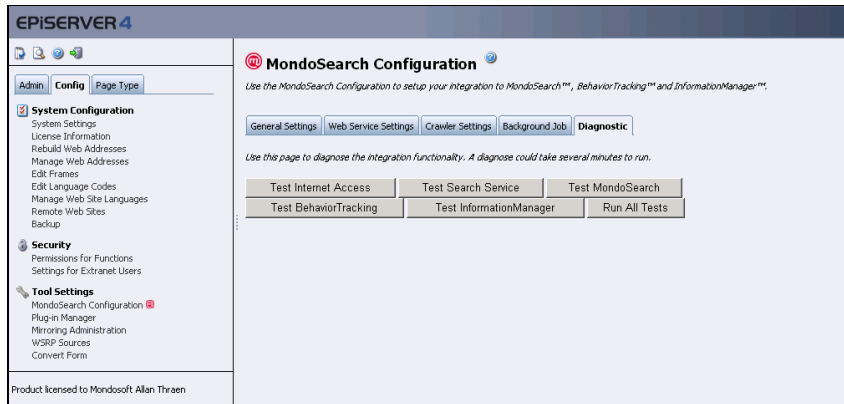
## Using the Diagnostics Tool

The **MondoSearch Configuration** page includes a diagnostics tool that checks to make sure that EPiServer is able to contact the Internet and all of the MondoSearch Web services it needs. It also checks to make sure that the remote services are “alive”. Among other things, this can help you make sure that you have entered your Web service URLs and connection strings correctly (see also “Configuring Web Service Communications” on page 49).

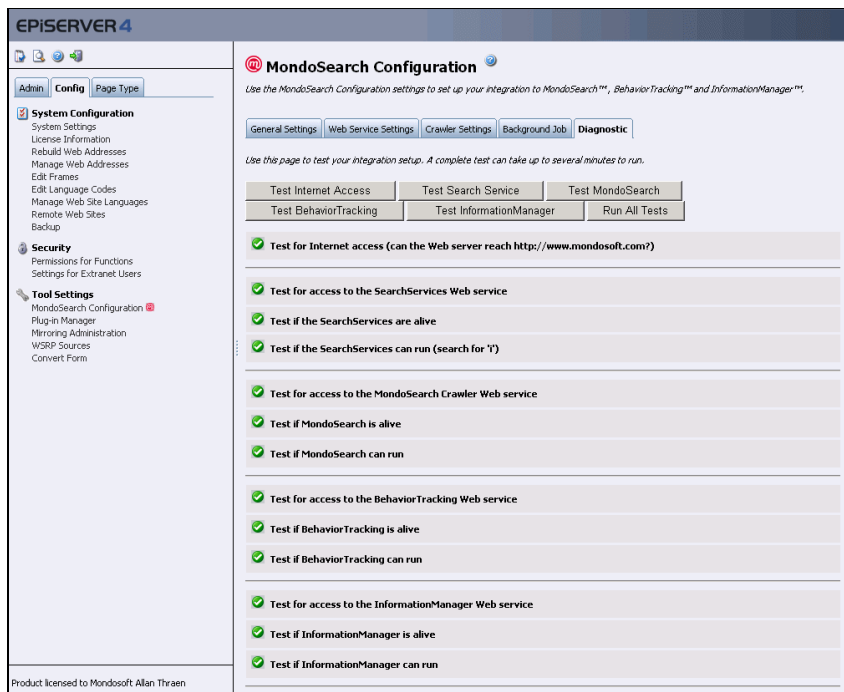
To run the diagnostics test:

1. Go to the **MondoSearch Configuration** page and select the **Diagnostics** tab (see also “Accessing the MondoSearch Administration Settings” on page 46).





2. Click on one of the buttons provided to run the test named (or click on **Run All Tests** to run all of them). Several options are provided here because it can sometimes take a few minutes to run all of the tests on a large installation.
3. The diagnostic tool runs its set of tests and shows the result of each of them. If a problem is indicated, check the relevant settings and try again. If problems persist, then you may need to check with your hosting-service provider or inspect your local MondoSearch application installations.







# CHAPTER 6

## Manual Crawler Control and Monitoring

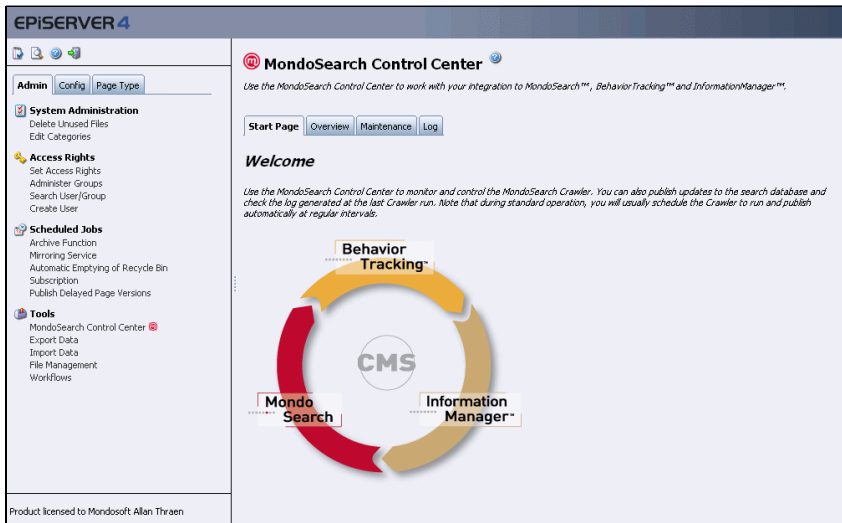


The MondoSearch for EPiServer package installs MondoSearch Crawler control and monitoring tools into the EPiServer administration interface. From here, you can issue commands to manually stop or start the Crawler. You can also monitor the Crawler status and view the log generated by the last Crawler run.

## Accessing the MondoSearch Control Center

Crawler control and monitoring tools are available in the **MondoSearch Control Center**. To access the Control Center:

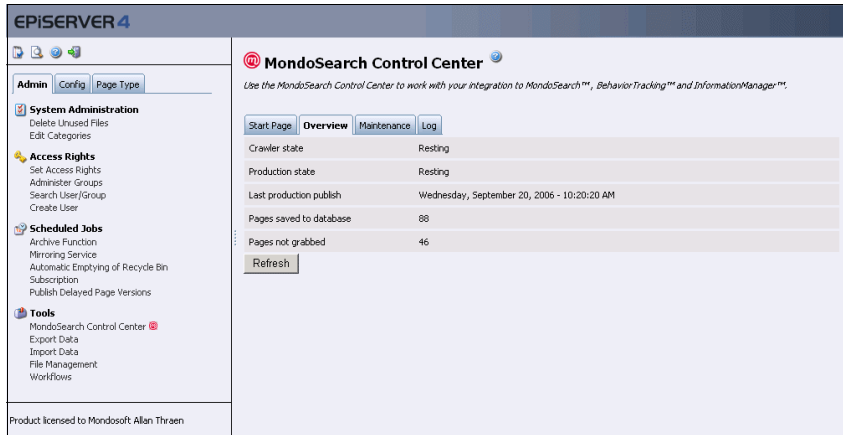
1. Log in to EPiServer as a user that has administrator privileges.
2. Enter Admin Mode (e.g., by using the right-click context menu or clicking on the **Admin** button while in Edit Mode).
3. Click on the **Admin** tab in the left frame (though this is usually selected already by default).
4. Click on the **MondoSearch Control Center** link under the **Tools** heading in the left frame.
5. When you first come to the **MondoSearch Control Center**, a welcome note is displayed. Use the various other tabs of the **MondoSearch Control Center** page in the right frame to navigate to the various displays and controls. All of the available functions are described in the remaining sections of this chapter.



## Crawler Monitoring and Status

The **MondoSearch Control Center** includes an overview of the current state of the MondoSearch Crawler and its integration with EPiServer.

To view the overview, go to the **MondoSearch Control Center** and select the **Overview** tab (see also “Accessing the MondoSearch Control Center” on page 60).



The following information is shown here:

- **Crawler state:** The current state of the Crawler (running, resting, failed, etc.)
- **Production state:** The current state of the Publish operation, which normally follows Crawling, but can also be executed separately (running, resting, failed, etc.)
- **Last production publish:** The date when the search database was last completely updated and then “published” to the production database, where it is available for searching by site users.
- **Pages saved to database:** The total number of pages currently listed in the production database.
- **Pages not grabbed:** The total number of pages that were found on the last crawl, but were not added to the database for some reason (such as limits on page size (too few words) or robot commands (do not crawl)). This number *does not* include pages that were not found (e.g., because they are not linked to from the rest of the site, they were behind pages that included a robot command of “do not link” or because some or all of the site did not allow Crawler access).

The information shown on this page is fetched when you first open the page, but is not refreshed while the page is open. If you want to update the display, then click on the **Refresh** button.

## Manual Crawler Control and Publishing

After first installing the integration package and/or at any time after making an especially important update to your site, you can choose to start the crawler manually to create/update the database. Also, if the Crawler is running and you discover that there is some kind of error or that there is too much load on your servers, then you can choose to kill the process (instant, but leaves the grab database in an incomplete and undefined state so you should *not* publish it).

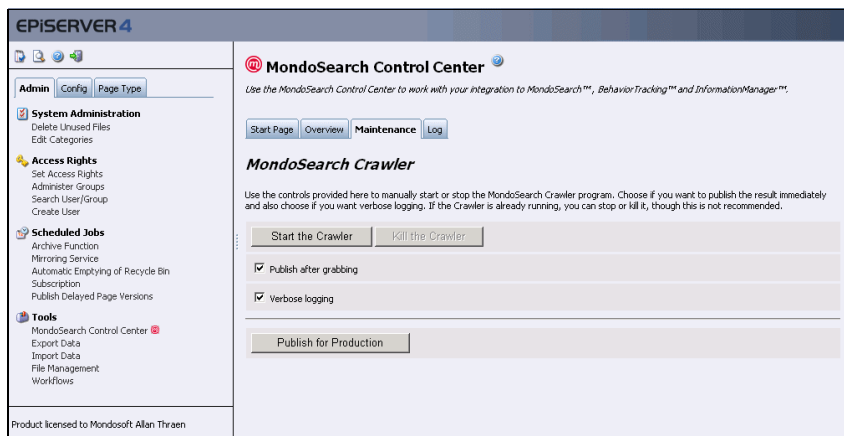
*Publishing*, in MondoSearch, refers to a process that usually occurs immediately after crawling. During crawling, the Crawler fills a database called the *grab database* (sometimes also called the *preview database*), which is independent from the *search database* normally used by users on your site. When crawling is finished, MondoSearch then processes the database, making optimizations and applying settings made in InformationManager. Finally, the grab database is copied to the search database; during the brief period of time it takes to make this copy, the search database is unavailable. When you issue a manual publish command, MondoSearch re-optimizes the grab database (again, applying, among other things, InformationManager features such as SearchHeaders) and copies the result to the search database.

There are basically two situations in which you might manually publish:

- After making a crawl in which you did not automatically publish (e.g., because you wanted to test against the grab database without affecting the search database)
- After you have made InformationManager settings that you want to apply right away without crawling.

To issue manual commands to the Crawler and/or Publisher:

1. Go to the **MondoSearch Control Center** and select the **Maintenance** tab (see also “Accessing the MondoSearch Control Center” on page 60).



2. If you are planning to run the Crawler then set the following options as needed:

- **Publish the User Interface after Grabbing:** Mark this box to make the new database available immediately after crawling. Clear this box to update the grab (preview) database only; this can be useful for testing new settings without affecting your public search database. See your MondoSearch user and developer documentation for more information about these two database and how to make use of the grab database during testing.
- **Verbose Logging:** if you would like to generate an extra-detailed log of Crawler actions then mark this check box (this can be useful when troubleshooting the Crawler). Clear this box to generate the standard level of log detail, which requires fewer resources and gives a quicker overview.

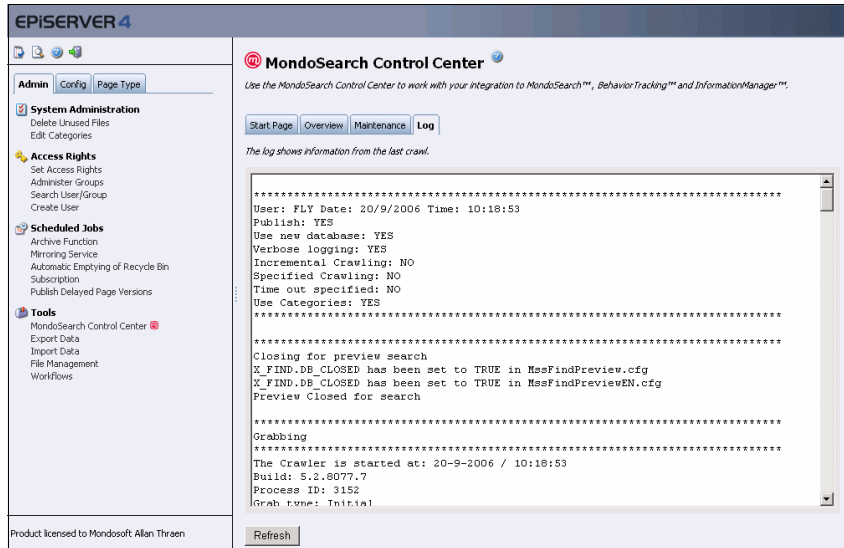
These settings do not affect scheduled Crawler runs.

3. Depending on what you want the Crawler to do, click on one of the following buttons:
  - **Start the Crawler:** Click here to begin crawling right away with the options selected in the check boxes described above. If the button is grayed out, then the Crawler is already running.
  - **Kill the Crawler:** Click here to stop the Crawler instantly. This leaves the grab database in an incomplete and undefined state so you should *not* publish it. If the button is grayed out, then the Crawler is not running.
  - **Publish for Production:** Click here to re-apply InformationManager settings to the current grab database and copy it to the search database, where it will be available to all users of your site. Usually this is done automatically after each crawl, but you might use this button if you have been testing without publishing, or if you have made new InformationManager settings (such as SearchHeaders) that you would like to apply right away without crawling. See also Chapter 12: “InformationManager Features in EPiServer”.

## Reading the Crawler Log

Each time the Crawler runs, it generates a log of its activities. If you are having trouble with the Crawler, this is a good place to look to find out what is going wrong.

To view the log, go to the **MondoSearch Control Center** and select the **Log** tab (see also “Accessing the MondoSearch Control Center” on page 60).



If the Crawler runs again while this page is open, you can fetch the new log by clicking on the **Refresh** button.





# CHAPTER 7

## Search/Result Page Configuration in EPiServer



The MondoSearch search/result page contains controls that enable it to generate either a search form, result list or both (depending on its current state—if the page receives a query, then it shows results; otherwise it shows a complete search form).

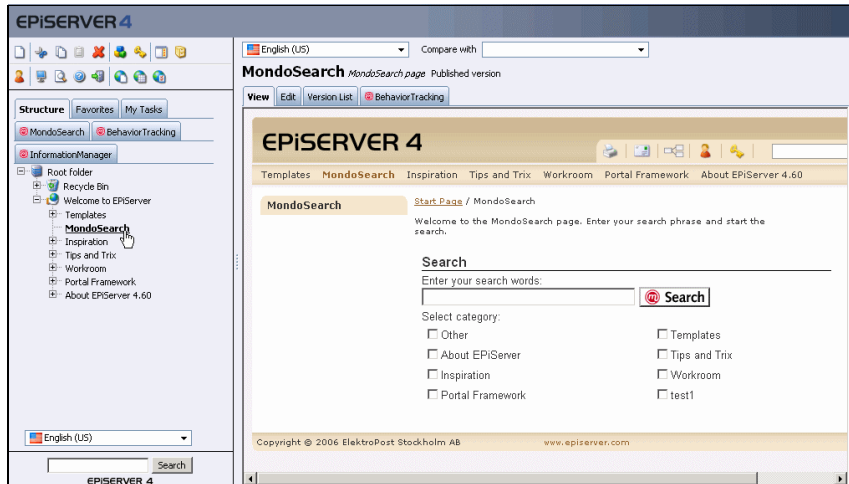
Often, your site will include just one page of this type, which will be the target of all of your simple search fields and also the target of an “advanced search” link. However, you might add several instances of the MondoSearch page, each configured with different options to perform a different type of search based on who the user is, where he or she is on your site or other criteria.

The default MondoSearch page type installed by the integration package provides a collection of controls for making some of the most commonly used search settings. You can further refine the page’s behavior by specifying search options using the Mondosoft MQL query language or by editing the page and its underlying code in Microsoft Visual Studio, but this is usually not necessary.

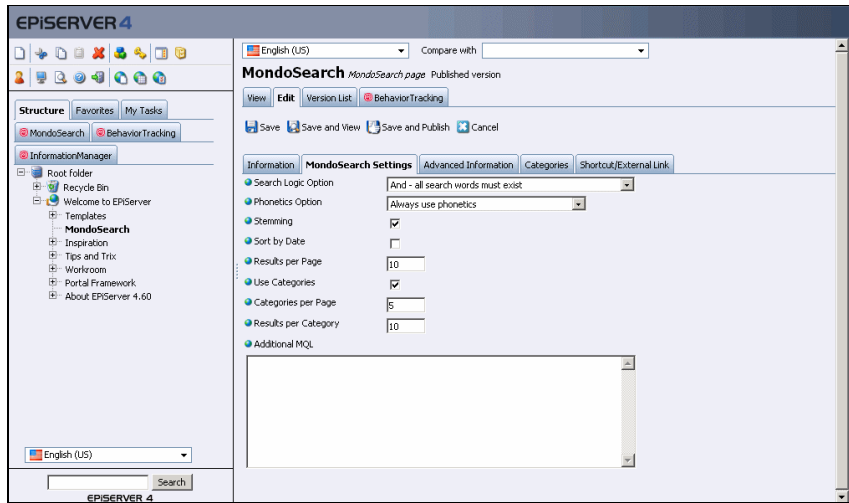
## Configuring Search Options

To set the standard options for a MondoSearch search/result page in your site:

1. Log in to EPiServer as a user that has editing privileges.
2. Enter Edit Mode (e.g., by using the right-click context menu).
3. Click on the **Structure** tab in the left frame.
4. Navigate to and select the MondoSearch for which you want to specify search options (or create a new page of the MondoSearch type). A preview of the page is then shown in the right frame.



- Click on the **Edit** tab in the right frame. This takes you to the edit view for your selected page. This view itself has several sub-tabs. Click on the **MondoSearch Settings** sub-tab.



Here, you will find the following settings, where are provided by the MondoSearch page type:

- Search Logic Option:** controls the type of logic applied to multi-word queries. Select “Auto-Boolean (Or)” to apply logic that finds pages that include at least one of the words, with pages that include all or most of the words ranked highest (recommended). Select “And” to return only pages that include all of the words (often generates short result lists). Select “Exact” to return only pages that include the exact phrase entered by the user (generates very short lists). Note that users can also choose to make an exact-phrase search at any time (regardless of the setting here) by enclosing the query in quotes. There is also an empty (blank) selection, which will use whatever setting is currently configured in MondoSearch (you can set this using the MondoSearch InSite, its native configuration interface).
- Phonetics Option:** controls whether and how MondoSearch should apply its phonetic-match algorithm. This algorithm helps compensate for bad spelling by finding words that “sound like” the search words specified by the user. When enabled, this option normally generates longer result lists. Select “Always apply phonetics” to apply phonetics to all searches. Select “Use phonetics when no results are found” to limit the algorithm to affect only searches that would otherwise produce empty results. Select “Never use phonetics” to disable the algorithm. Select “Default” to use whatever setting is currently configured in MondoSearch (you can set this using the MondoSearch InSite, its native configuration interface).
- Stemming:** enables the MondoSearch stemming algorithm. When stemming is enabled, the search engine will return results that include words that are gram-

matically related to the submitted search word(s). For example, if the user searches for “running puppy”, then MondoSearch will also find pages that include the words “run”, “ran”, “runs” and/or “puppies” (unless the user searches for an exact phrase match). Queries will often return many more results when stemming is enabled.

- **Sort by Date:** mark this check box to have MondoSearch give extra weight to recently added pages. Usually, this means that newest pages will appear at the top of the list, but other factors will still be considered, such as keyword matches and results of auto-boolean logic.
  - **Categories per Page:** sets the maximum number of categories that will be shown on a multi-category page (see also Figure 1 on page 12). If more than this number of categories is found, then the user will need to paging links to step through the full list categories.
  - **Results per Page:** sets the maximum number of documents that will be listed on single-category pages (see also Figure 2 on page 13). If more than this number of documents is found for a given category, then the user must click on paging links to step through the full list of results.
  - **Use Categories:** enables the MondoSearch categories system (highly recommended). When this box is cleared, MondoSearch categories will not be shown in the search form or result lists, but existing assignments will be remembered and you can still make category assignments, which will take effect if you later enable this check box. See Chapter 8: “Setting up MondoSearch Categories” for more information about the categories feature.
  - **Results per Category:** sets the maximum number of documents that will be listed under each category on the multi-category page (see also Figure 1 on page 12). If more than this number of documents is found for a given category, then the user will need to click on a link to see the full list of documents for that category.
  - **Additional MQL:** the MondoSearch SearchService Web service is extremely flexible. As mentioned earlier, the controls on the MondoSearch page communicate with the SearchService using the Mondosoft Query Language (MQL), which is extremely flexible and powerful. Using MQL, you can make many settings that affect the content, ranking and format of the returned result set. The settings described above are all examples of this; when you make those settings, you are actually just modifying the MQL generated by the MondoSoft search controls. Any text that you add in the **Additional MQL** will be appended to the MQL generated by the MondoSearch controls; because it comes last, your custom code will overrule the automatically generated code if any conflicts occur. For complete details about the MQL syntax, please see the *MondoSearch Developer's Guide*. See also Appendix B: “Advanced Integration and Development” in this manual.
6. When you are done making settings, click on one of the save buttons at the top of the **Edit** tab to apply your settings. See also your EPiServer documentation for advice about how to use these buttons.

## Graphic Design and Advanced Customization

Further graphic customization is possible by editing the search/result page directly, for example using Microsoft Visual Studio., though some .NET programming experience is probably required to do so effectively. In fact, you can modify all aspects of the appearance and functionality of the page using standard .NET techniques. In a standard installation, the search/result page is installed as:

---

```
<EPiServerSiteRoot>\templates\units\MondoSearchPage.aspx
```

---

Where *<EPiServerSiteRoot>* is the root content folder for your site.





# CHAPTER 8

## Setting up MondoSearch Categories



MondoSearch normally displays search results that are organized into categories. This often helps users find the pages they are looking for more quickly. For this feature to function, you must establish the categories you wish to use and assign your various pages to them as needed. This chapter explains how the categorization feature works and how to create and assign categories within EPiServer.

## Planning Your Categories

Before you begin to categorize your site, start by analyzing the full scope of your site and its contents. Usually, you should try to come up with 4 to 10 categories into which all of your pages can be divided. There must be enough categories to ensure that search results will be suitable divided, but not so many that users have trouble choosing the right one in which to search. Usually, you will choose categories that roughly match the site structure that you already have in place; because MondoSearch categories can be inherited along site structure lines, this is also the easiest way to assign categories to many pages at once.

The following rules apply for MondoSearch categories:

- Each page belongs to *exactly one* MondoSearch category.
- You can have as many MondoSearch categories as you want. Usually, the MondoSearch Crawler will automatically create a new category for each unique category value it finds while crawling the site.
- MondoSearch categories are totally independent of EPiServer categories described in the EPiServer documentation; the two systems have no effect on each other.
- All pages placed below a given categorized page in the site structure will inherit the category of that senior page unless overridden by a subsequent category assignment.
- New pages added to a branch of your structure for which a MondoSearch category has already been assigned will automatically inherit that category.
- If more than one category is inherited by a given page, then the closest category will be the one that applies to the page. For example, if both the parent and grandparent of a given page have categories applied, then the category of the parent will override the grandparent. A given page's own category assignment overrides all inherited assignments and will furthermore apply to all current and future child pages unless overridden further down in the structure.
- If you delete the category assigned to a specific page, then that page will revert to inheriting the category of one of nearest ancestor page.

## How MondoSearch Categorization Works

The MondoSearch integration package includes a control that generates several types of META-tags, which provide extra page information that is indexed by the MondoSearch Crawler and made searchable in MondoSearch (see also Chapter 3: “The META-Tag Emitter”). Among the META-tags generated by the control is one that identifies the MondoSearch category for the page.

The system tries to obtain the categories from two different sources as follows:

1. If a MondoSearch category is established in the dynamic properties as described in “Making Category Assignments” on page 73, then this is always used. This category then becomes the value of the CATEGORY META-tag generated for the page.
2. If no MondoSearch category is established, then the system checks to see if one or more standard EPiServer categories have been assigned (see your EPiServer documen-



tation for details about this system. If so, then the first category listed is used as the MondoSearch category. This category then becomes the value of the CATEGORY META-tag generated for the page.

3. If still no category assignment is found, then the page is indexed as having no category. The CATEGORY META-tag is therefore not added to the page at all.

Note that the MondoSearch META-tag generator control is usually configured to deliver its enhanced META-tags only to the MondoSearch Crawler and not to other types of Web crawlers or standard browsers. See “Setting the Crawler Agent ID” on page 55 for details about how to configure this option.

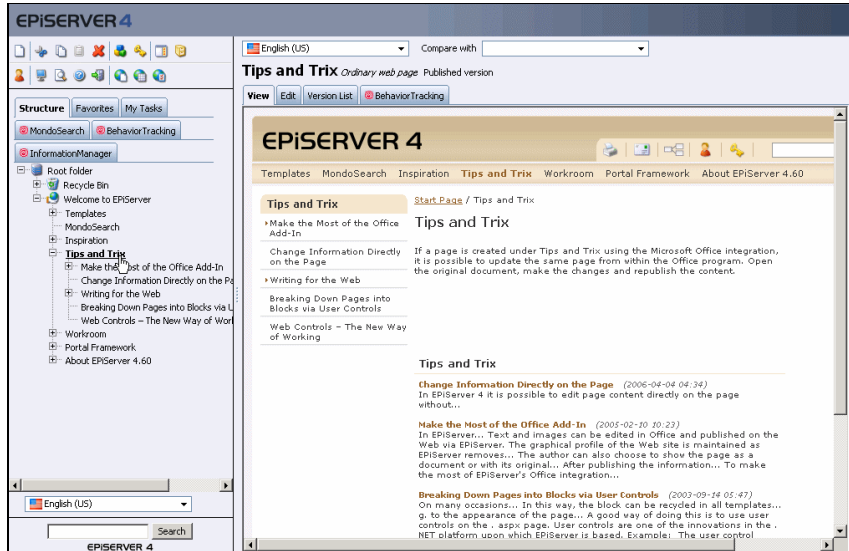
Note also that as an alternative, or supplementary, system of applying categories, you could use the CatMap feature of the MondoSearch InSite. Please see the *MondoSearch User's Guide* for details.

Finally, note that non-HTML documents, such as PDF or Microsoft Office files, do not provide META-tags for categorization. If your site includes these types of documents, then please see “Categorizing Binary Document Types” on page 75 for details about how to categorize them.

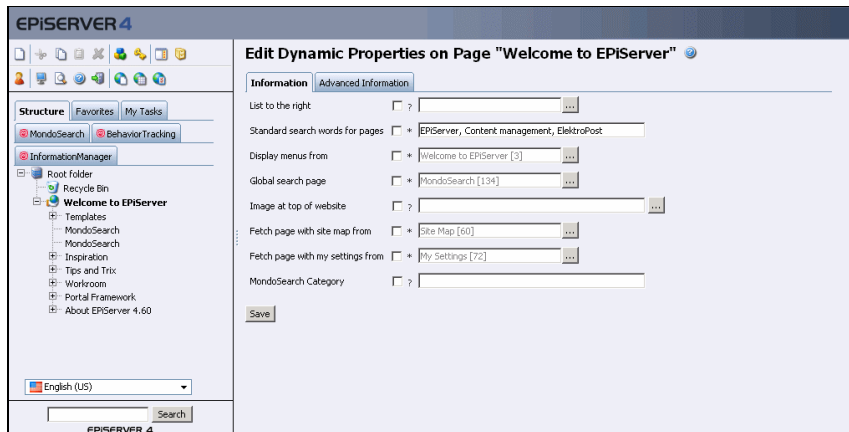
## Making Category Assignments

To assign a MondoSearch category to a page and/or branch of your site structure, do the following:

1. Log in to EPiServer as a user that has editing privileges.
2. Enter Edit Mode (e.g., by using the right-click context menu).
3. Click on the **Structure** tab in the left frame.
4. Navigate to and select the page for which you want to specify a MondoSearch category (or create a new page). A preview of the page is then shown in the right frame.



5. Click on the **Edit dynamic properties** button in the edit toolbar. The button looks like three colored cubes stacked into a pyramid.
6. The dynamic properties view opens in the right frame. Most of the settings here are part of the standard EPiServer installation; however, there is also now a **MondoSearch Category** setting, which is added by the integration package (please see your EPiServer documentation for more information about how to work with dynamic properties).



The **MondoSearch Category** setting includes a check box, a symbol (\*, ? or ^) and a text-input field. They work as follows:

- **Input field:** Enter a category name in this field to specify a category for the current page (and all of its child pages that have no MondoSearch category setting). Only one category can be assigned, but the category can include more than one

word. Usually, the MondoSearch Crawler will automatically create a new category for each unique name it finds as it crawls your site. This makes it very easy to create a new category, but it also means that you must be careful with your spelling. To set the current page to inherit its category, clear the text from this field (or leave it blank).

- **Check box:** To clear all categories assigned to all children (and grandchildren, etc.) of the current page, mark the check box. The result will be that all children of the current page will revert to having no local MondoSearch category setting, thereby causing them to inherit their category from your current page.
- **Status symbol:** As is standard for dynamic properties in EPiServer, this symbol tells you how the page currently receives its MondoSearch category setting. The “^” symbol indicates that the page is inheriting its category; hold your mouse pointer over the symbol to read the value it is inheriting and the page ID from which it comes. The “\*” symbol indicates that the value is set explicitly for the current page (therefore, the current value is shown in the text field). The “?” symbol indicates that the page has no category assigned.

Enter a category in the field provided (and, if desired, mark the check box to overrule any existing categories in child pages). Then click on **Save** to apply the setting.

## Categorizing Binary Document Types

In addition to HTML pages, MondoSearch can also index the full content of PDF, Microsoft Office and Adobe Flash files. However, because these types of documents do not supply META-tags, like HTML documents do, they do not support the META-tag system outlined in “How MondoSearch Categorization Works” on page 72.

If your site includes PDF and/or MS Office documents, then EPiServer usually saves these in a single location, regardless of the areas of your site where the documents are available (via links). Unfortunately, this collects documents of all categories into the same folder, which makes them difficult to categorize. One simple way to handle this is to use the MondoSearch CatMap to assign categories to individual documents, establish wildcards or apply a single, generic category (e.g., “Resources”) to all of the documents in your upload folder. (Please see the *MondoSearch User's Guide* for details about how to use the CatMap) Alternatively, you can modify your MondoSearch configuration so that each binary document will inherit the category of the HTML file that links to it.

If you are using a hosted version of MondoSearch, then please contact your service provider to request the “categorize by referring link” feature. If you are running a local MondoSearch installation, then you can enable this feature as follows:

1. Find the following file from your MondoSearch installation:

---

```
<SearchHostInstallFolder>\data\msscscat.cfg
```

---

Where *<SearchHostInstallFolder>* is the folder in which you have installed your search host files (e.g., *C:\MondoSearch\www.MySearchHost.com*).

2. Edit the *msscat.cfg* file with a text editor and add the following line:

---

```
X_CAT.CLICKPATH="<ResourceFolderPath>"
```

---

Where *<ResourceFolderPath>* is the path to the upload folder where EPiServer keeps the binary documents that you upload (often, this is *"/upload/"*).

This setting causes MondoSearch to categorize all files found under the named path according to the category of the URL that links to them.

3. Save your changes.



# CHAPTER 9

## Providing Authorized Result Lists



MondoSearch is able to ensure that result lists only include links to documents that the current user has permission to view. This makes result lists easier to use and helps enhance security by ensuring that potentially revealing document titles and descriptions are not shown to unauthorized users. This chapter explains how the authorization system works and how to configure it to meet your needs.

## Introduction

On many sites, especially intranets and those that provide both public and private areas, the content available to a given user varies according to that user's access level. For these sites, the Web server will automatically refuse to deliver restricted documents to unauthorized users; however, the search engine uses a separate database to generate its result sets, which means that search results could potentially list documents that the current user will not be able to load. This is not only annoying to the user, but can also create a security breach if the document titles and/or descriptions are revealing enough. MondoSearch therefore provides several mechanisms for providing result sets that include only documents to which the current user has access. These include:

- *Result-set filtering*  
When this system is in place, MondoSearch keeps a list of user groups that have access to each page and filters the result list to include only those pages that the searching user will be able to see. Using this method, all paging controls and result counts will be accurate and sensitive documents will be totally hidden. In most cases, we recommend this technique if it is possible for your site. Note, however, that all non-HTML documents will not have META-tags and, therefore, will be hidden if you use this method.
- *Document suppression*  
When this system is in place, the MondoSearch result page checks, at search time, each found document with the Web server to see whether the current user is able to access that URL. To conserve resources, it checks only those documents scheduled to be listed on the result page—not the full result set. Unauthorized documents are simply left off the page. However, the paging controls and result counts will all be based on the full result set, so the perceptive user might find this information confusing. One advantage of this technique is that your security settings will apply to result lists as soon as they are changed on the Web server.
- *Limited-access indicator*  
This system is similar to the document-suppression system except that, instead of hiding the unauthorized documents, they are displayed but grayed out to indicate that they cannot be loaded. This results in accurate result counts and paging controls, but risks exposing sensitive information if it appears in page titles and/or deceptions.

The strengths and weaknesses of the various methods are summarized below in Table 2. Note that you can combine the methods as required.

Further details and instructions for implementing of these techniques are given in the remaining sections of this chapter.

Method	Advantages	Disadvantages
Result-set filtering	<ul style="list-style-type: none"> <li>• Very fast and resource efficient</li> <li>• Accurate result counts</li> <li>• Sensitive pages are never shown</li> </ul>	<ul style="list-style-type: none"> <li>• Security settings are only updated as often as the Crawler indexes your site. This means that if you change the access level of a given document, then that document will still appear in result lists to unauthorized users until the next crawl. However, your Web server will not deliver the full document.</li> <li>• This approach only works for HTML files. Other types of documents (PDF, Office, etc.), will not include the required META-tags and will therefore be removed by the filter so your results will never include these types of documents.</li> </ul>
Document suppression	<ul style="list-style-type: none"> <li>• Document-access changes made on the Web server apply immediately to search results</li> <li>• Automatically applies to all types of documents</li> <li>• Sensitive pages are never shown</li> </ul>	<ul style="list-style-type: none"> <li>• This technique is applied <i>after</i> the result list has been generated and checks are made only for the documents scheduled to appear on the current result page. This means that result counts (e.g., total number of pages found) will be inaccurate. It also means that result pages may be shorter than expected; for example, if the page should show ten results, but three fail the authorization test, then the page will only show seven results.</li> <li>• Places extra demands on the Web server because each listed page must be tested against the Web server at search time.</li> </ul>
Limited-access indicator	<ul style="list-style-type: none"> <li>• Document-access changes made on the Web server apply immediately to search results</li> <li>• Automatically applies to all types of documents</li> <li>• Accurate result counts are preserved</li> </ul>	<ul style="list-style-type: none"> <li>• Users will be able to read the titles of sensitive pages, which may sometimes represent a security breach. However, you could also design the feature to hide these.</li> <li>• Places extra demands on the Web server because each listed document must be tested against the Web server at search time.</li> </ul>

**Table 2** Strengths and weaknesses of the various methods of authentication result lists

## Configuring the Authentication System

The search/result page supplied with the integration package is prepared to apply any combination of the authentication systems described in this chapter. The functionality is controlled by a few parameters whose values are defined near the top of the page's code. To edit the features for a given search/result page, you must open the file in a text editor and edit the values of these parameters as required.

The following settings control the authentication feature:

- `EPiServerSecurityFilter="True" | "False"`  
Set this to “true” to enable the result-set filtering method (recommended). Set to false to ignore the authentication information stored in the search database.
- `EPiServerSecurityRenderCheck="True" | "False"`  
Set this to “true” to enable the document-suppression method. Set to “false” to list all documents that passed the result-set filter without checking each link with the server.
- `EPiServerSecurityShowMessage="True" | "False"`  
Set this to “true” (when document suppression is being used) to configure the system to show suppressed documents but indicate that they are not accessible. Set to “false” to hide suppressed documents (can result in inaccurate result counts).

For example:

To use pure result-set filtering as outlined in Table 2, use the following settings:

- `EPiServerSecurityFilter="True"`
- `EPiServerSecurityRenderCheck="False"`
- `EPiServerSecurityShowMessage="False"`

To use pure document suppression as outlined in Table 2, use the following settings:

- `EPiServerSecurityFilter="False"`
- `EPiServerSecurityRenderCheck="True"`
- `EPiServerSecurityShowMessage="False"`

To use document suppression with limited-access indicators outlined in Table 2, use the following settings:

- `EPiServerSecurityFilter="False"`
- `EPiServerSecurityRenderCheck="True"`
- `EPiServerSecurityShowMessage="True"`





# CHAPTER 10

## Editor Search



The integration package installs a new search interface for page editors and authors. It is available via the new **MondoSearch** tab added to the EPiServer Edit Mode interface. Unlike the built-in editor search functionality of EPiServer (which searches only for page names and IDs), the MondoSearch feature indexes the full *rendered* versions of your pages, just as the site visitor sees them. The MondoSearch feature returns a ranked result list from which editors can select a page, which opens for immediate editing.

## Editor Search Appearance and Availability

By default, Editor Search is available to all users in Edit Mode. However, you can configure the solution to restrict access to this feature to a specific user group or to disable the feature altogether. You can also choose to show all of the MondoSearch tabs in Edit Mode using short titles; in this case, the Editor Search will be indicated with **MS** rather than **MondoSearch** (as shown in the procedure in the next section).

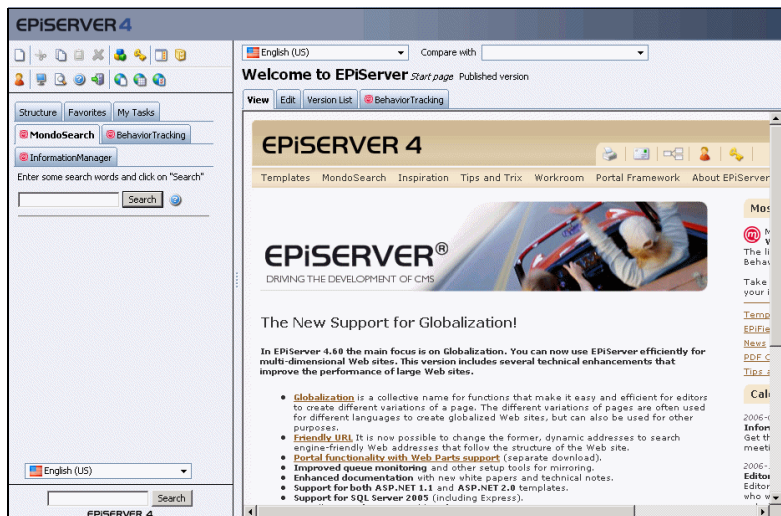
To change the appearance and availability of the Editor Search feature, you must be able to enter Admin Mode in EPiServer. See “Controlling Edit-Mode Features’ Appearance and Availability” on page 46 for details about how to make the relevant settings.

If you would like to use Editor Search feature, but cannot see it and do not have access to Admin Mode, then please contact your administrator for help.

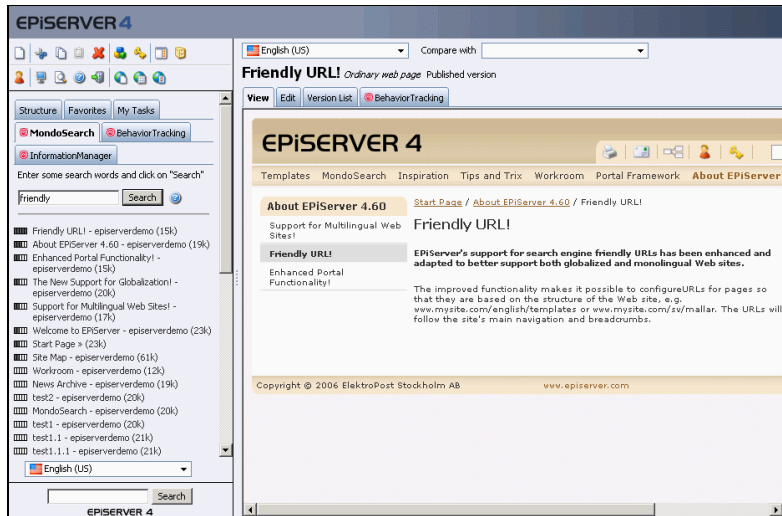
## Using the Editor Search

To run the Editor Search:

1. Log in to EPiServer as a user that has editing privileges.
2. Enter Edit Mode (e.g., by using the right-click context menu).
3. Click on the **MondoSearch** tab in the left frame. This opens a search form.



4. Enter a query and click on **Search**. A list of matching pages is now shown in the left frame, ranked by relevance.



5. Click on a result in the list to open the target page in the edit area in the right frame.





# CHAPTER 11

## Working with BehaviorTracking in EPiServer



BehaviorTracking provides advanced search-log analysis tools that enable authors, designers and managers to evaluate how well the search engine is meeting the needs of its users, to see what users are looking for and to decide how best to improve the site content and search engine settings.

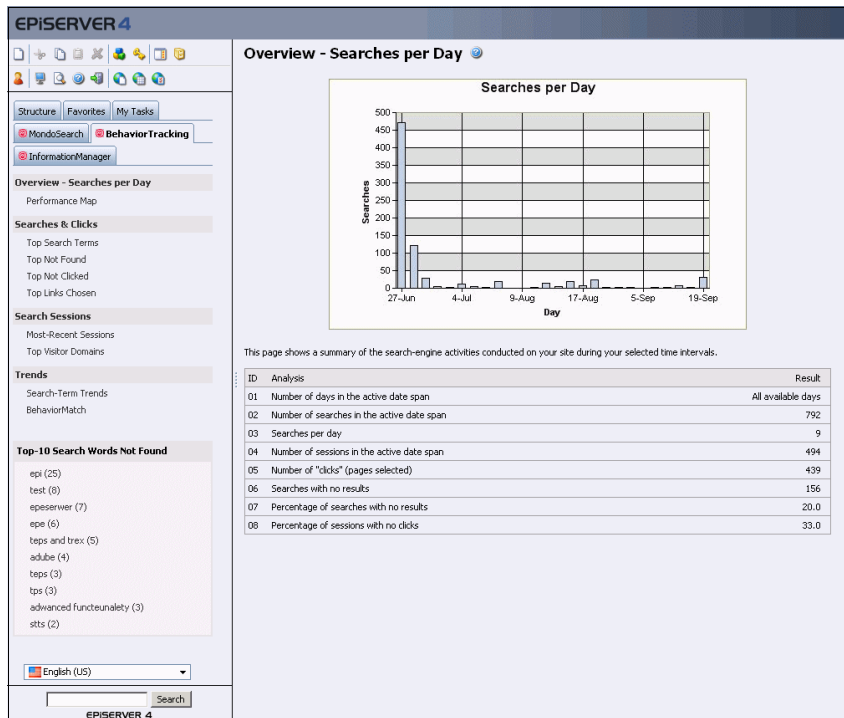
This chapter describes how to find and interpret all of the many BehaviorTracking views added to EPiServer by the integration package. See also “Detailed Logging and Reporting with BehaviorTracking” on page 13 for a detailed overview of BehaviorTracking.

## Finding BehaviorTracking in EPiServer

To view BehaviorTracking data in EPiServer, do the following:

1. Log in to EPiServer as a user that has editing privileges.
2. Enter Edit Mode (e.g., by using the right-click context menu).
3. Click on the **BehaviorTracking** tab in the left frame.
4. Use the various links under the **BehaviorTracking** tab to view the various BehaviorTracking tools described in the remaining sections of this chapter.

For more information about logging in to EPiServer and entering Edit Mode, please see your EPiServer documentation.



Below the navigation links, the **BehaviorTracking** tab shows a list of the top ten search words for which no results were found. This is one of the most useful lists generated by BehaviorTracking because it indicates some of the most common problems people are having when trying to find things on your site. The list names each of the not-found terms and shows (in parenthesis) the number of sessions in which each listed term was used. For more information about a term listed here, click on the target term. This list is the same as the one given on the Top Not-Found Terms page; see "Top Not-Found" on page 94 for more advice about how to make use of this information. See also "Term

Details” on page 103 for details about how to interpret the information shown when you click on one of these terms.

**Note:** Though the **BehaviorTracking** tab is provided by default, EPiServer administrators might choose to hide it. If you wish to use BehaviorTracking, but cannot see the tab while in Edit Mode, then please talk with your EPiServer administrator. See also “Controlling Edit-Mode Features’ Appearance and Availability” on page 46.

## The BehaviorTracking Date Span Setting

All of the BehaviorTracking statistics shown in EPiServer is based on search-log data collected over a specific time frame called the *date span*. You can configure the date span to include the past 7, 14, 60, 90 or 365 days; or choose to view statistics based on all collected data. You (or your site administrator) should select this value based on the nature of your site, the number of users you have and what you are trying to find out.

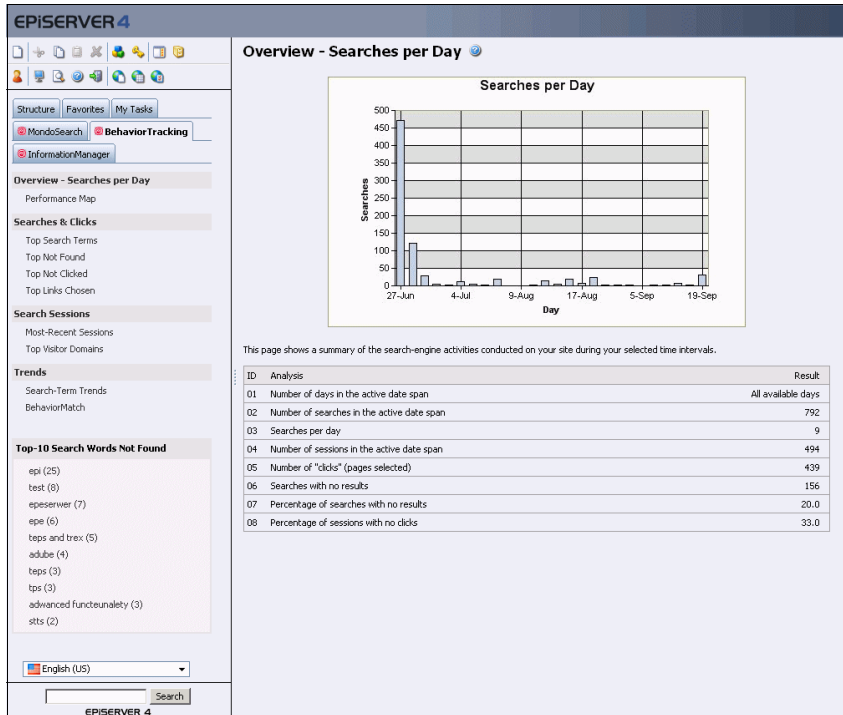
See “Setting the BehaviorTracking Date Span” on page 48 for details about how to configure the date span.

## BehaviorTracking Overview

### Overview—Searches per Day

The **Overview - Searches per day** page shows a summary of all of the search-engine activities conducted on your site over the current date span.

To view this page, go to Edit Mode in EPiServer, click on the **BehaviorTracking** tab in the left frame and then click on the **Overview - Searches per day** link.



The information given on this page gives you the broadest overview of how people are using your search engine. This information may help guide you to other areas of BehaviorTracking, thereby enabling you to improve both your Web site and your search-engine configuration. Each of the values shown here is described below with notes about how to interpret the readings.

### Reading the Graph

The graph at the top of this page shows how busy the search engine has been each day over the current date span.



## Reading the Statistics

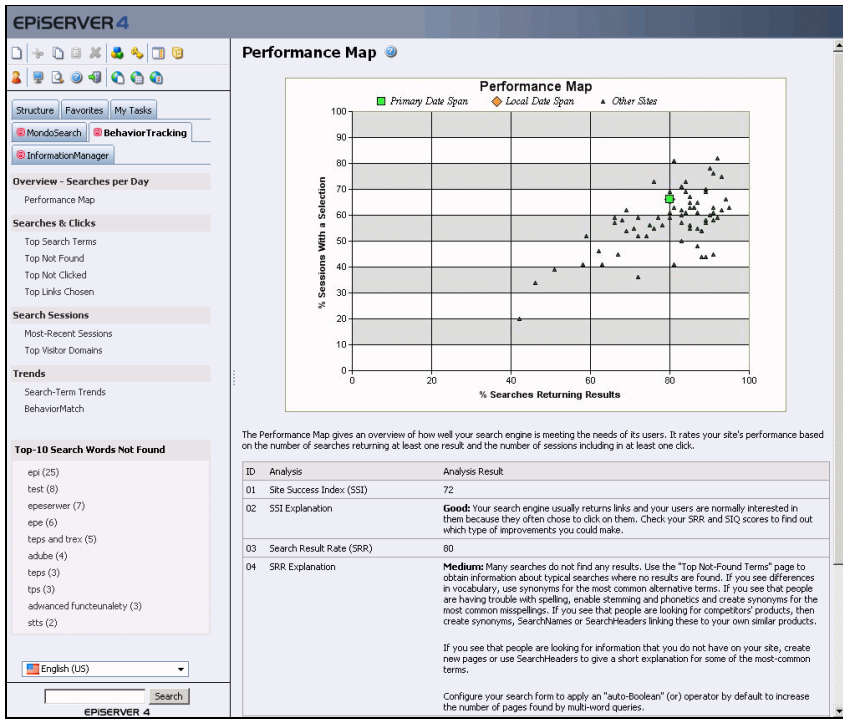
At the bottom of the **Overview** page are listed the following statistics:

- **Number of days in the active date span:** Shows the number of days in the date span currently being considered by BehaviorTracking. All of the statistics shown in BehaviorTracking are based on data gathered from searches during this period. See “Setting the BehaviorTracking Date Span” on page 48 for details about how to configure the date span.
- **Number of searches in the active date span:** This value shows the total number of times that a search was made within the date span. BehaviorTracking counts each unique search term that occurs in a single session as a single search, regardless of how many times that term is re-submitted during the session. Therefore, actions such as clicking on a paging button to step through multi-page search results and clicking on a “show all results in category” link are counted as being part of the same single “search”. Note, however, that search terms are sensitive to word order, so “red apple” would be considered a different search from “apple red”, even when occurring in the same session.
- **Searches per day:** The average number of searches made each day.
- **Number of sessions in the active date span:** This value indicates how many times somebody used the search engine. BehaviorTracking considers all search requests coming from a single IP address with no more than ten minutes break between requests to belong a single session. You should consider a session to represent a single full attempt by a user to find a specific piece of information.
- **Number of “clicks” (pages selected):** In BehaviorTracking, the term “click” refers to the action of a user clicking on a link from the search-engine result page. This value indicates the total number of clicks that occurred over each time span.
- **Searches with no results:** This indicates the number of times that a user submitted a search term for which no pages were found.
- **Percent of searches with no results:** This indicates the proportion of searches where no pages were found as a percentage of the total number of searches.
- **Percentage of sessions with no clicks:** This value indicates the proportion of sessions where users never loaded any of the pages listed on the result page as a percentage of the total number of sessions.

## Performance Map

The Performance Map gives an overview of how well your search engine is meeting the needs of its users. It rates your site's performance based on the number of searches returning at least one result and the number of sessions including at least one click.

To view this page, go to Edit Mode in EPiServer, click on the **BehaviorTracking** tab in the left frame and then click on the **Performance Map** link under the **Overview** heading.



## Interpreting the Graph

At the top of the **Performance Map** page, a graph is displayed showing how your site performs based on the following measures:

- **% Searches returning a result (result rate)**  
Only searches that return at least one matching page have any chance of being useful. Searches where no pages were found are considered to indicate a failure. If your result rate is high, then your search engine is working well. If a large percentage of searches are not returning any results, then your users may be misunderstanding the purpose of your site, or your search engine might be too restrictive; consider enabling features such as auto-boolean logic, stemming and synonyms.
- **% Sessions with a selection (success rate)**  
A "click" is registered whenever a user loads one of the pages listed in the result list. A session with at least one click is considered to indicate a success in the eyes of the user. If your click percentage is close to your result percentage, then your users are probably pleased with the results they see.

The graph shows the position of your site during each of the following two date spans:

- **Primary:** This is the date span configured in the native BehaviorTracking interface. Usually, this includes all available time, but you could change this. See the *BehaviorTracking User's Guide* for details.
- **Local:** This is the date span currently configured in EPiServer. See “Setting the BehaviorTracking Date Span” on page 48 for details about how to configure this.

Sample points for other typical sites are also plotted; these are static and based on the large number of sites that use Mondosoft's hosting service. Use these additional points to gauge how your site ranks among other sites that use MondoSearch. Generally, the closer your site comes to the upper-right corner, the better its performance is.

### Interpreting the Advice Table

Below the graph is a table that gives some interpretations of your site's performance. The interpretations are generated based on your site's position along the two axis of the graph and on its score for the following two index values:

- **Site Success Index (SSI)**  
 The site success index (SSI) indicates how closely your site matches the expectations of its visitors. It is calculated based on your site's absolute distance from the “optimal” score in which both percentages on the graph are at 100 (the upper right corner). The closer your site is to this optimum, the higher its success index will be. SSI is reported as a percentage value between 0 and 100 (with 100 being the best). It is computed as shown in the equation below, where *SCR* is the percentage of sessions with at least one click and *SRR* is the percentage of searches with at least one result.

$$SSI = 100 - \frac{\sqrt{(100 - SRR)^2 + (100 - SCR)^2}}{\sqrt{2}}$$

- **Search Implementation Quality (SIQ)**  
 This indicates how often users select a link after submitting a search that returned at least one result. It shows the implementation's ability to make people interested in the results after they have searched. Generally, users only get a chance to select a link when at least one link was provided, so the best efficiency is the one where the percentage of sessions with a click is the same as the percentage of searches that return at least one result. The SIQ, therefore, is essentially the vertical distance between your site's position and a diagonal line drawn from the bottom left to the upper right corner. This is just a rough estimate because users might conduct more than one search in a session (especially when the first search returned no results). However, our research indicates that users rarely submit more than one or two searches, regardless of how satisfied they are with the results. Though your SIQ can sometimes rise over 100, this will rarely be the case unless you are calculating with a very small amount of data; values close to 100 are considered to indicate an optimal score, with lower values indicating some level of user disinterest in the search results. SIQ is calculated as shown below, where *SCR* is the percentage of sessions with at least one click and *SRR* is the percentage of searches with at least one result.

$$SIQ = 100 - (SRR - SCR)$$

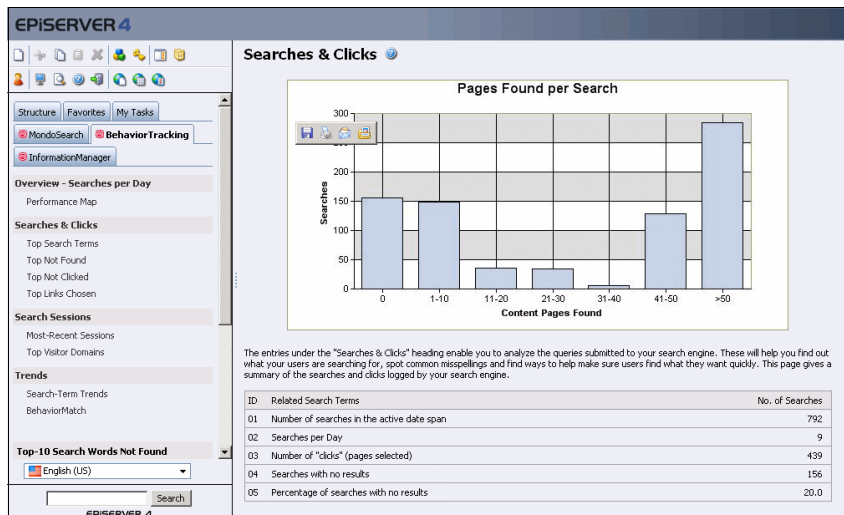
In addition to these raw scores, the table also provides interpretations of your scores and provides some advice for how to react. These interpretations are based on the results found for the large number of diverse, real-life sites that use the Mondosoft search-hosting service. The explanations are very general and alternative explanations are always possible, so you should consider both the graph and the interpretations to be simple guidelines.

## Searches & Clicks

### Searches & Clicks Overview

The **Searches & Clicks** page gives you a few statistics and a graph to provide an overview of the search queries submitted to your site.

To view this page, go to Edit Mode in EPiServer, click on the **BehaviorTracking** tab in the left frame and then click on the **Searches & Clicks** link.



The graph at the top of this page shows how many times the search engine returned a given range of found pages. Statistics below this graph reenforce these figures.

- **Number of searches in the active date span:** this is the total number of times a user submitted a *new search term* during a session. Actions such as clicking on a paging button on multi-page search results and clicking on a "show all categories" link on the result page are counted as being part of the same search. This value is equal to the sum of each bar in the graph.
- **Searches per day:** this is the average number of searches that were made per day.
- **Number of "clicks" (pages selected):** this shows the total number of times that any user clicked on a link listed on the result page from your search engine.

- **Searches with no results:** this indicates the number of times a user submitted a query for which no pages were found.
- **Percentage of searches with no results:** This is calculated based on the values for **Searches with no results** and **Number of searches**.

See also “Acting on BehaviorTracking Data” on page 108 for advice for how to react to some of the results presented here.

## Top Search Terms

The **Top Search Terms** page shows a list of the terms that users submitted most often to the search engine and how many times each of them was submitted over the specified time span. Terms are ranked with the most-used at the top of the list.

To view this page, go to Edit Mode in EPiServer, click on the **BehaviorTracking** tab in the left frame and then click on the **Top Search Terms** link.

**EPISERVER 4**

**Top Search Terms**

This page lists the search terms most-often submitted by users of your site. Click on a term for more details about that term.

ID	Search Terms	Sessions
01	epi-server	72
02	news	50
03	epi	25
04	news archive	22
05	tips	20
06	blogs	19
07	tips and tricks	18
08	new epi-server	17
09	pdf documents	14
10	publish pages	13

Number of items: 10 Refresh

As you study the list, consider the following:

- Terms at the top of this list describe the things that most visitors to your site are trying to find. Consider establishing SearchNames for these or adding links on your home page that lead directly to these topics. See also “SearchNames” on page 116.
- When planning to upgrade or expand your site, check this list to find out what types of things people are most interested in finding.

To view more details about any of the terms listed on this page, click on the target term. This will bring you to the **Term Details** page, which will show a list of related terms, sessions in which the term was used and the pages that were selected after searching for that term. See “Term Details” on page 103 for more information.

## Top Not-Found

The **Top Not-Found Terms** page shows a list of the search terms for which no documents were found. The total number of sessions in which each term was used is also shown. The terms are ranked with the most-used at the top of the list.

Note that the list shown on the **Top Not-Found Terms** page is the same as the list shown in the left frame while the **BehaviorTracking** tab is open.

To view this page, go to Edit Mode in EPiServer, click on the **BehaviorTracking** tab in the left frame and then click on the **Top Not Found** link.

The screenshot shows the EPiServer 4 BehaviorTracking interface. The left sidebar contains navigation links: Structure, Favorites, My Tasks, MondoSearch, BehaviorTracking (selected), and InformationManager. Under BehaviorTracking, there are sections for Overview - Searches per Day, Searches & Clicks (with links to Top Search Terms, Top Not Found, Top Not Clicked, and Top Links Chosen), Search Sessions, and Trends. The main content area is titled 'Top Not-Found Terms' and includes a descriptive text: 'This page lists the most often-used search terms that did not return any results. Consider configuring your search engine with synonyms to correct common misspellings and to match your products to generic terms and/or the names of similar competitors' products.' Below this is a table with two columns: ID and Sessions. The table lists 10 search terms and their corresponding session counts. At the bottom of the table, there is a 'Number of items' dropdown set to 10 and a 'Refresh' button.

ID	Search Terms Not Found	Sessions
01	epl	25
02	test	8
03	epiServer	7
04	epe	6
05	teps and trex	5
06	adube	4
07	teps	3
08	tps	3
09	advanced functionality	3
10	stts	2

As you study the list, consider the following:

- Very often, you will find several misspellings in this list. Consider establishing common misspellings as synonyms to the correct word or adding them to the keyword META-tags for pages that are likely to help. You could also add a list of pre-defined search-links for those popular searches that many users have difficulty spelling. Put it right under the search form.
- Some users may be using different words than you or your page authors normally do (e.g., help/support, car/vehicle/automobile, address/location). As with misspellings, consider setting up synonyms in InformationManager or add common synonyms to keyword META-tags for the appropriate pages.
- If users are searching for old products, consider adding these as synonyms in InformationManager or to the keyword META-tags for pages describing products that have now replaced these old products.
- If there are numerous searches for products that you do don't carry, consider adding these products to your inventory, or making a page explaining that you do not carry these products, maybe even with a link to a site that does (especially if it is not a competitor). This will be considered a very good service by your users.

- Your site might be missing the information that users are seeking. Consider adding information about topics that many users seem to be interested in.

Though not currently part of the standard integration package, InformationManager enables you to establish synonyms within MondoSearch. You can still take advantage of this feature for your EPiServer site, but you must use the native InformationManager interface to configure your synonyms; see the *InformationManager User's Guide* for details.

You may see an entry called “Empty search”. This refers to cases when a user has clicked on the submit button without entering a search term. (If your site places default text into the search field (e.g., “type your query here”), then this may appear instead of “Empty search”.) There are several reasons why this may happen:

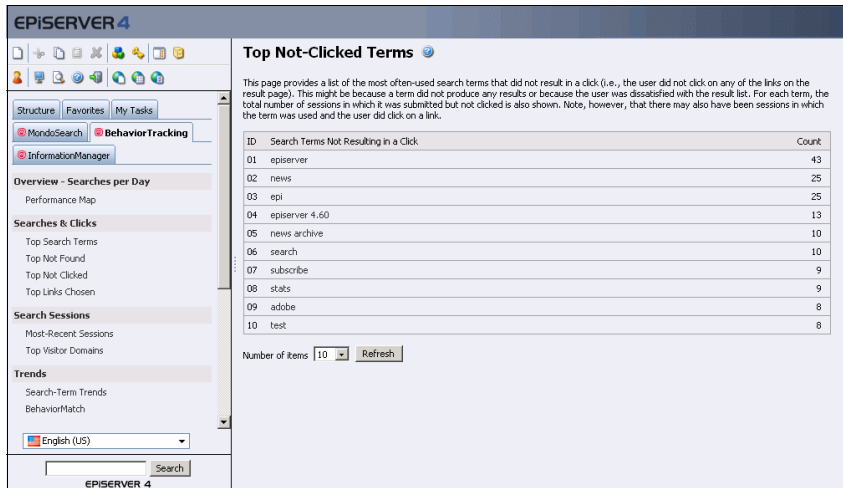
- The users may not recognize the search field and are clicking on the submit button because they think it will lead to a search form. This most typically happens when you have a simple input field and search button on your home page, and then most typically when you are using a graphic rather than a standard button.
- If you are using a non-standard label on your search button (e.g., “Tell me” or “Do it!”), then confused users might click on it without realizing that they should enter a query first.
- Users may be pressing the return key on their keyboard without realizing that this will submit the search form.
- Web crawlers might be following the search button as a link, thereby submitting an empty form. Consider modifying your page(s) so that the Crawler will not submit the search form; see “Marking Content for Exclusion from the Search Database” on page 43 for advice.

To view more details about any of the terms listed on this page, click on the target term. This will bring you to the **Term Detail** page, which will show a list of related terms, sessions in which the term was used and the pages that were selected after searching for that term. See “Term Details” on page 103 for more information.

## Top Not-Clicked Terms

The **Top Not-Clicked Terms** page provides a list of the most often-used search terms that did not result in a click (i.e., the user did not click on any of the links on the result page). This might be because a term did not produce any results or because the user was dissatisfied with the result list. For each term, the total number of sessions in which it was submitted but not clicked is also shown. Note, however, that there may also have been sessions in which the term was used and the user *did* click on a link. The terms are ranked with the most-used at the top of the list.

To view this page, go to Edit Mode in EPiServer, click on the **BehaviorTracking** tab in the left frame and then click on the **Top Not Clicked** link.



ID	Search Terms Not Resulting in a Click	Count
01	epi-server	43
02	news	25
03	epi	25
04	epi-server 4.60	13
05	news archive	10
06	search	10
07	subscribe	9
08	stats	9
09	adobe	8
10	test	8

Number of items: 10 Refresh

There are several reasons why a user might not follow any of the links he or she has found:

- The desired information was on pages that were ranked too low, so the user did not page down far enough to find them. Consider using keyword, rank-word and/or rank-document META-tags to fix this. See also the *MondoSearch User's Guide* for details about how to control page rankings using META-tags.
- The user was not able to identify the correct page because the titles and/or descriptions are vague, misleading, or the same for many pages. See “Optimizing Page Titles and Descriptions” on page 42 for guidelines about how best to do this in EPiServer.
- Your site does not contain the information that the user was looking for—a situation made clear by the search results the user saw. Depending on the search word, this might not be a problem. However, if users’ intentions are clear, consider adding information to your site that will help users searching for the most popular words listed on the **Top Not-Clicked Terms** page.
- The category with the correct link did not appear on the first page of the multiple-category list. We have found that 19 out of 20 users do not go to the second multiple-category page because they assume that the best results will appear on the first page. Usually, MondoSearch will place the categories containing the best matches first; however, it is possible to customize the solution to use fixed-position categories, which means that category position will not indicate best results. Consider returning to ranked categories or increasing the number of categories displayed on the result page so that all categories appear on one page. (Switching to fixed-position categories requires that you add custom MQL to your search form and that you configure category positions in the MondoSearch InSite; see *The MondoSearch User's Guide* for details about the InSite and see the *MondoSearch Developer's Guide* for details about writing MQL; see “Configuring Search Options” on page 66 for details about adding MQL to your search forms and controlling the number of categories per page.)

To view more details about any of the terms listed on this page, click on the target term. This will bring you to the **Term Detail** page, which will show a list of related terms, sessions

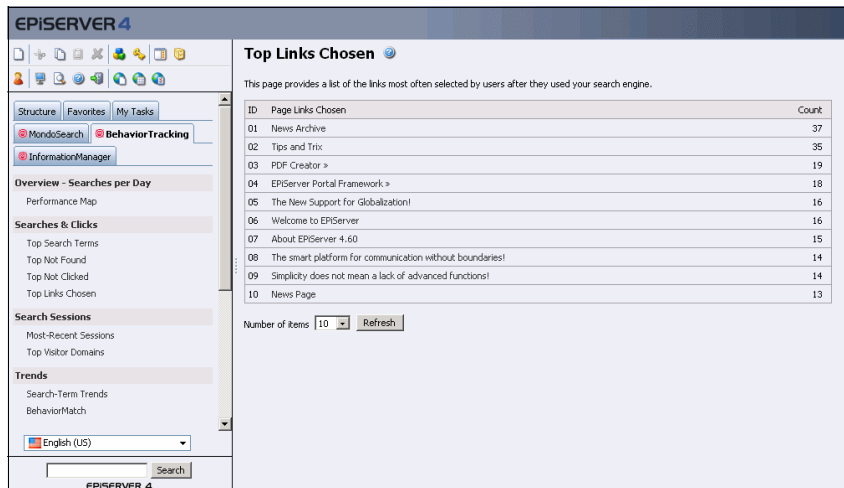


in which the term was used and the pages that were selected after searching for that term. See “Term Details” on page 103 for more information.

## Top Links Chosen

The **Top Links Chosen** page shows a list of the pages most often selected by users after they used your search form. For each page, the URL and the number of times that a link to that page was followed are shown.

To view this page, go to Edit Mode in EPiServer, click on the **BehaviorTracking** tab in the left frame and then click on the **Top Links Chosen** link.



The screenshot shows the EPiServer 4 BehaviorTracking interface. The left sidebar contains navigation tabs: Structure, Favorites, My Tasks, MondoSearch, BehaviorTracking (selected), and InformationManager. Under BehaviorTracking, there are sections for Overview - Searches per Day, Searches & Clicks (with links to Top Search Terms, Top Not Found, Top Not Clicked, and Top Links Chosen), Search Sessions, and Trends. The main content area is titled 'Top Links Chosen' and includes a description: 'This page provides a list of the links most often selected by users after they used your search engine.' Below this is a table with 10 rows, each showing an ID, a page link, and a count. At the bottom of the table, there is a 'Number of items' dropdown set to 10 and a 'Refresh' button.

ID	Page Links Chosen	Count
01	News Archive	37
02	Tips and Tricks	35
03	PDF Creator »	19
04	EPiServer Portal Framework »	18
05	The New Support for Globalization!	16
06	Welcome to EPiServer	16
07	About EPiServer 4.60	15
08	The smart platform for communication without boundaries!	14
09	Simplicity does not mean a lack of advanced functions!	14
10	News Page	13

The list on this page gives you an idea of which pages of your site are most popular. You might consider making some of these pages even easier to find by making SearchNames or adding links for them from your home page if appropriate.

You might also find that a page has been selected *too* often when another page would have been better, possibly because it has been ranking too highly or because it has a deceptive title or description. You can improve this, for example, by creating better page titles and descriptions or by allowing your search engine to generate unique titles and descriptions automatically. You might also use META-tags to add keywords or ranking commands that might help users find the correct page more easily. See Chapter 4: “Optimizing Your Site for Search” for more information about these and other techniques.

To see complete details about one of the pages listed, click on the target page in the list. See “Page Details” on page 106 for more information about how to interpret the page details.

## Search Sessions

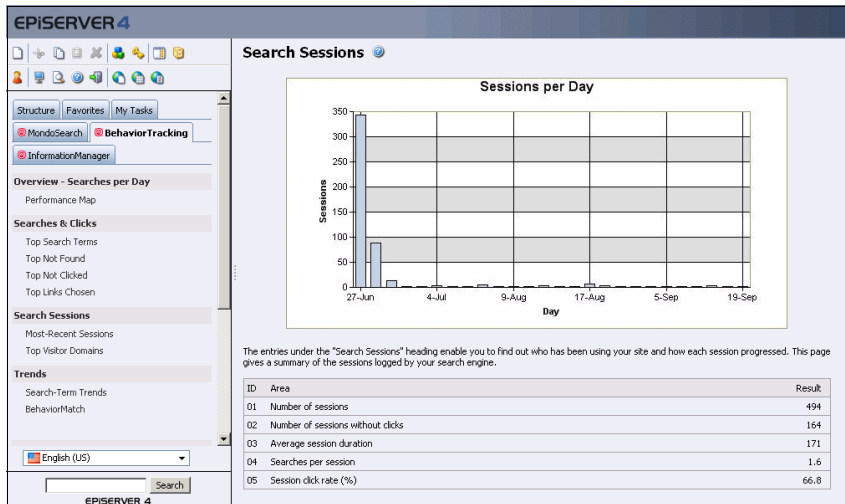
BehaviorTracking considers all search requests coming from a single IP address with no more than ten minutes break between requests to belong a single session. You should consider a session to represent a full attempt by a user to find a specific piece of information.

The **Search Sessions** area of BehaviorTracking enables you to analyze the various sessions registered with your search engine to find out who has been using your site and how each session progressed.

### Search Sessions Overview

The **Search Sessions** page gives you few statistics and a graph to provide an overview of the search sessions that have occurred on your site.

To view this page, go to Edit Mode in EPiServer, click on the **BehaviorTracking** tab in the left frame and then click on the **Search Sessions** link.



The graph at the top of this page shows how many sessions occurred on each day of the current date span. The table at the bottom of the page summarizes the session information found.

- **Number of sessions:** Shows the total number of sessions recorded. This value indicates how many times somebody used the search engine during the specified time span. This will give you an idea of how popular your search engine is. If this value is lower than you expected, then users might be having trouble finding your search form. Consider making an easy-to-find link on your home page, or even including a simple search form there.
- **Number of sessions without clicks:** Shows the total number of sessions during which the user did not load any of the pages found by the search engine.

- **Average session duration:** Shows the average of how long (in seconds) each session lasted.
- **Searches per session:** Shows how many searches each user typically made during a session. Numbers between 1 and 3 are normal. Generally, a low value here indicates that users are quickly finding what they are looking for. Remember that BehaviorTracking considers all searches for a given term during a single session to be a single “search”; this means that actions such as paging and switching between categories will be counted as part of the same search so long as the user does not change the query.
- **Session click rate (%):** This is calculated based on the values for “number of sessions” and “number of sessions without clicks”. Generally, a value close to 100 indicates that users are finding pages that look interesting on your result page. However, if this is combined with a large number of searches per session, then user might not be happy with the pages they are finding after loading them and therefore return to use the search engine again.

## Most-Recent Sessions

The **Most-Recent Sessions** page lists the most-recent sessions by IP number or (when available) the domain name of the searcher.

To view this page, go to Edit Mode in EPiServer, click on the **BehaviorTracking** tab in the left frame and then click on the **Most-Recent Sessions** link.

**EPISERVER 4**

**Most-Recent Sessions**

This page lists the most-recent user sessions. Each session is identified either by the IP number or (when available) the domain name of the searcher.

ID	Session	Date
01	megan.vvm.net	9/19/2006
02	cpshserver08.mondosoft.local	9/19/2006
03	192.168.2.50	9/18/2006
04	192.168.2.50	9/18/2006
05	cpshserver08.mondosoft.local	9/15/2006
06	cpshserver08.mondosoft.local	9/15/2006
07	cpshserver08.mondosoft.local	9/15/2006
08	cpshserver08.mondosoft.local	9/15/2006
09	diportal00.msdinet.local	9/14/2006
10	diportal00.msdinet.local	9/14/2006

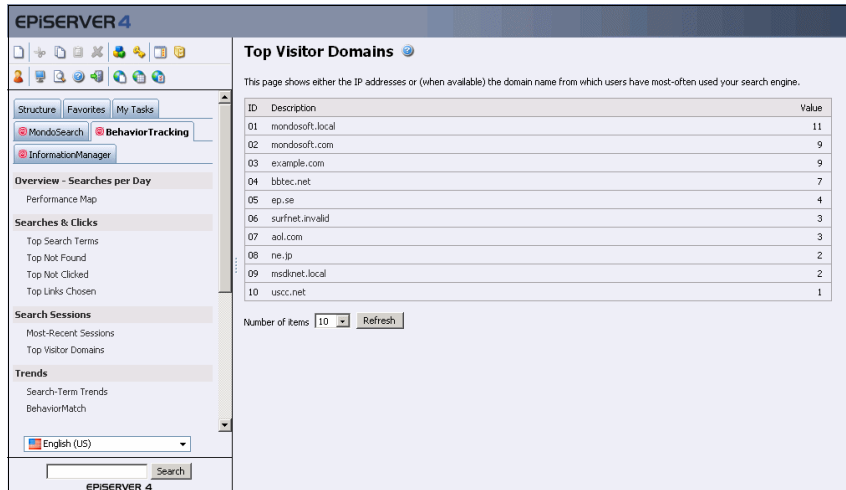
Number of items: 10 Refresh

To view complete details about a listed session, click on its name in the **Session** column. See “Session Details” on page 105 for more information about that page.

## Top Visitor Domains

The **Top Visitor Domains** page shows the IP addresses and (when available) domain names for the users that have most used your search engine and the number of sessions originating from each of them.

To view this page, go to Edit Mode in EPiServer, click on the **BehaviorTracking** tab in the left frame and then click on the **Top Visitor Domains** link.



**EPISERVER 4**

**Top Visitor Domains**

This page shows either the IP addresses or (when available) the domain name from which users have most-often used your search engine.

ID	Description	Value
01	mondosoft.local	11
02	mondosoft.com	9
03	example.com	9
04	bbtec.net	7
05	ep.se	4
06	surfnet.invalid	3
07	aol.com	3
08	ne.jp	2
09	msdinet.local	2
10	uscc.net	1

Number of items:

Click on a link in the **Description** column to view a list of sessions from that domain.

Consider the following as you study this list:

- Look for interesting companies, such as partners, investors, competitors, journalists or others. This will enable you to see what these interesting parties were looking for, which might help your company in future negotiations, marketing campaigns and other strategic planning.
- Look at the top-level domains (e.g., .com, .uk, .nl, .dk, etc.). These often indicate the country where the user lives. This might, for example, help you decide whether to support a new language on your site or decide where to concentrate your marketing efforts.
- You might notice that ISPs such as AOL or World Online often appear near the top of your list. Note that these services reuse IP numbers over and over again, assigning them at random to users as they come on and off line. Therefore, sessions from these domains are typically from many different (often private) users.

## Trends

The **Trends** area of BehaviorTracking provides a fast way to see how usage patterns have been changing on your search engine.

### Search-Term Trends

The **Search-Term Trends** page highlights the biggest changes in search-term usage among the search-engine users. It tracks the most-used terms and generates two tables: one that shows the terms whose popularity has most increased during the date span and one that shows the terms whose popularity has most decreased. Both of the tables here are sorted with the *fastest-moving* terms at the top.

To view this page, go to Edit Mode in EPiServer, click on the **BehaviorTracking** tab in the left frame and then click on either the **Trends** or **Search-Term Trends** link (both of these open the same **Search-Term Trends** page).

**EPISERVER 4**

**Search-Term Trends**

This page highlights the biggest changes in search-term usage on your site. It tracks the 100 most-used terms and generates two tables: one that shows the terms whose popularity has most increased during each date span and one that shows the terms whose popularity has most decreased. Both of the tables here are sorted with the fastest-moving terms at the top. Previous rankings are calculated for the periods immediately before the current date span and for the same number of days. Terms that were not in the top-100 during the previous period show a previous ranking of "-".

**Advancing Terms**

ID	Term	Current	Previous	Up
01	episerver	1	-	99
02	news	2	-	98
03	epl	3	-	97
04	news archive	4	-	96
05	tips	5	-	95
06	blogs	6	-	94
07	tips and trix	7	-	93
08	new episerver	8	-	92
09	pdf documents	9	-	91
10	publish pages	10	-	90

**Retreating Terms**

ID	Term	Current	Previous	Down
----	------	---------	----------	------

Number of items: 10 Refresh

Use the tables on this page to look for the effects of changes you have made to your search engine and/or Web site. You can also look here to see how new marketing campaigns, press releases, products and other external influences have affected what people are looking for on your site.

For each term listed in each table, the following information is provided:

- **Term:** this is the term as submitted by the users. Click on a term to view more information about it on the **Term Details** page (see also “Term Details” on page 71).
- **Current:** shows the popularity of the term during the date span. A value of 1 indicates the most popular term.
- **Previous:** shows how popular the term was previously. The previous ranking is calculated for the period immediately before the current date span and for the same number of days. For example, if the date span is set to cover the previous 30 days, the

previous ranking will be calculated based on the 30 days before that. Terms that were not in the top-100 during the previous period show a value of “-”.

- **Up / Down:** shows how far the term has moved up or down in popularity (it's previous ranking minus its current ranking). Terms that were not in the top-100 at the previous ranking are calculated as though they ranked at exactly 100 previously.

## BehaviorMatch

BehaviorTracking is able to analyze usage patterns to discover common search terms that were unsuccessful on a first search attempt, but which did succeed when changed slightly, for example to correct spelling or use a synonym. This feature is called *BehaviorMatch*. These misspellings and synonyms are listed on the **BehaviorMatch** page of BehaviorTracking.

To see the list of BehaviorMatch terms and suggested synonyms, go to Edit Mode in EPiServer, click on the **BehaviorTracking** tab in the left frame and then click on the **BehaviorMatch** link.

ID	Original Search term	Suggested synonyms
01	adobe	pdf documents
02	adube	pdf documents
03	epe	epeserver
04	epeserver	epe
05	epi	episerver, episerver 4.60
06	episerver	epi, episerver 4.60, search, test
07	episerver 4.60	epi, episerver
08	news	news archive, news archive
09	news archive	news
10	news archive	news
11	pdf documents	adobe
12	pdf documents	adube
13	search	episerver, test
14	statistics	stats
15	stats	statistics
16	subscribe	subscription
17	subscription	subscribe
18	tips	tips

You can use the list shown here to establish META-tags and synonyms that match the common misspellings and alternative terms to the appropriate pages of your site. See your EPiServer documentation for details about how to create key-word META-tags for pages.

Note that another way to add META-tags to your pages based on BehaviorTracking data is to use the single-page BehaviorTracking tab; see “Page Details” on page 106 for details.

Note that the stand-alone InformationManager application enables you to create synonyms automatically from BehaviorMatch suggestions, but you must manually approve each one before it will take effect. See the *MondoSoft InformationManager User's Guide* for more information about how to do this.

## Term Details

The **Term Details** page provides detailed information about a specific search term. Details include:

- A list of related terms
- A list of sessions in which the term was used
- A list of pages that were selected after searching for that term

The **Term Details** page appears whenever you select a term from one of the other pages of BehaviorTracking that lists search terms. Because a term must already be selected, you cannot navigate directly to this page using the navigator.

The screenshot shows the EPiServer 4 interface. On the left is a sidebar with navigation links: Structure, Favorites, My Tasks, MondoSearch, BehaviorTracking (selected), and InformationManager. Below these are sections for Overview - Searches per Day, Searches & Clicks, Search Sessions, Trends, and Top-10 Search Words Not Found. The main content area is titled 'Term Details for: pdf documents'. It includes a description of the page's purpose, a 'Related Terms' table, a 'Most-Recent Sessions That Include This Term' table, and a 'Pages Chosen' table. At the bottom of the main area is a 'Number of Items' selector and a 'Refresh' button.

**Term Details for: pdf documents**

This page shows how a given search term was typically employed by users of your search engine. It lists other terms most-often used in sessions that included the current term, the most-recent sessions that include the current term and the links most-often selected after using the current term.

**Related Terms**

ID	Related Words	Sessions
01	adobe	8

**Most-Recent Sessions That Include This Term**

ID	Host Name	Date	Searches
01	101.46.82.21	6/27/2006	1
02	156.26.216.111	6/27/2006	1
03	106.196.70.13	6/27/2006	1
04	192.60.207.153	6/27/2006	1
05	151.225.94.178	6/27/2006	1
06	145.157.early-registration.of.surfnet.invalid	6/27/2006	1
07	107.220.223.192	6/27/2006	1
08	121.36.39.118	6/27/2006	1
09	176.117.223.104	6/27/2006	1
10	154.222.240.192	6/27/2006	1

**Pages Chosen**

ID	URL	Clicked
01	PDF Creator >	9
02	Inspiration	2
03	Site Map	1
04	Certificate	1
05	Search	1
06	Document List >	1

Number of Items: 10 Refresh

## Related Terms

Related terms are terms that were most-often submitted during sessions that also included the current term. These will help you get a better idea of the kinds of things people were trying to find when using the current term.

To see details about any of the terms listed, click on the target term. The **Term Details** page will then reload to give details about the new term.

## Most-Recent Sessions that Include this Term

This table lists the most-recent sessions in which the current term was used. The session list is sorted with the newest at the top.

To view complete details about a listed session, click on the host name for the target session in the list. By viewing the full details of a session, you may be able to figure out more precisely what a given person was looking for when he or she submitted the current term. See “Reading Session Details” on page 78 for more information about how to analyze sessions.

## Pages Chosen

This table lists the pages most often chosen after a search for the current search term. Use this list to help you decide if people seem to be finding the right pages. To see complete details about one of the pages listed, click on the target page in the list. See “Page Details” on page 106 for more information about how to interpret the page details.

If you think that people are being guided to the wrong page for a given word, then consider taking one or more of the following steps:

- Add the current term as a keyword or rank-word META-tag for the page (or pages) that you think are a better match.
- Establish the current term (or one of the words of the term) as a synonym for a word that appears often on pages that give a better match. (See the *BehaviorTracking User's Guide* for details about how to define synonyms.)



## Session Details

The **Session Details** page reports details about a single session, thereby allowing you to follow a single user's experience as though you were looking right over his or her shoulder.

To get to the **Session Details** page, you must click on a link on one of the other pages of BehaviorTracking on which sessions are listed.

**EPISERVER 4**

**Session Details (70277935)**

This page provides the click-by-click details of a single user session, thereby enabling you to follow a single user's experience as though you were looking right over his or her shoulder.

**Session Overview**

ID	Description	Value
01	Host name	101.46.82.21
02	Session date	6/27/2006
03	Session duration	0
04	Number of unique searches	1
05	Number of clicks	2
06	IP address	101.46.82.21

**Session Log**

ID	Time	Action	Word	Pages
01	10:16 PM	Search	pdf documents	57
02	10:16 PM	Search	pdf documents	0
03	10:16 PM	Search	pdf documents	0

Number of items: 10 Refresh

At the top of this page, you see summary information about the session, which specifies:

- **Host name:** Shows the domain name (if available) of the user who was active for the session. If no host name can be found, then the IP address will be shown instead.
- **Session date:** Shows the date when the session took place.
- **Session duration:** Shows how long the session lasted (in minutes).
- **Number of unique searches:** Shows the number of unique searches made during the session. All searches that use identical queries are considered the same unique search.
- **Number of clicks:** Shows the number of links that the user selected from the search-results page.
- **IP address:** Shows the IP address for the user who was active for the session.

At the bottom of the page is a complete list of events that occurred during the session. Two types of actions are logged:

- **Search:** For these entries, you will be able to read the number of pages found and the term submitted. To see details about a term, click on its link in the **Word** column; see “Term Details” on page 71 for more information about how to interpret term details.
- **Click:** The user clicked on a found page link. Often, sessions end here. Sessions that continue might indicate that the page was not what the user actually wanted. However, if the next search term is very different (especially if it occurs a few minutes later), then the user has simply thought of something else to look for already. To see additional details about a selected page, click on its link in the **Word** column; see

“Page Details” on page 69 for more information about how to interpret the page details.

All events are listed in sequence and indicate the time at which they occurred.

Sessions that show a large number of entries typically indicate one of the following:

- The user was having trouble finding the right information and therefore tried submitting many different queries. These are usually the most helpful type of session for finding out how to tune your search engine.
- The user was actually very successful, but was also very quick reading pages and then returning to search for something completely different in less than ten minutes. In some cases, these types of sessions may help you learn how to organize your information by clustering pages into areas of similar interest or appeal (for example with navigator links or search-result categories).

## Page Details

### Viewing Details for a Selected Page

The MondoSearch integration package enables you to view BehaviorTracking data collected for any individual page in your site. To view this data, do the following:

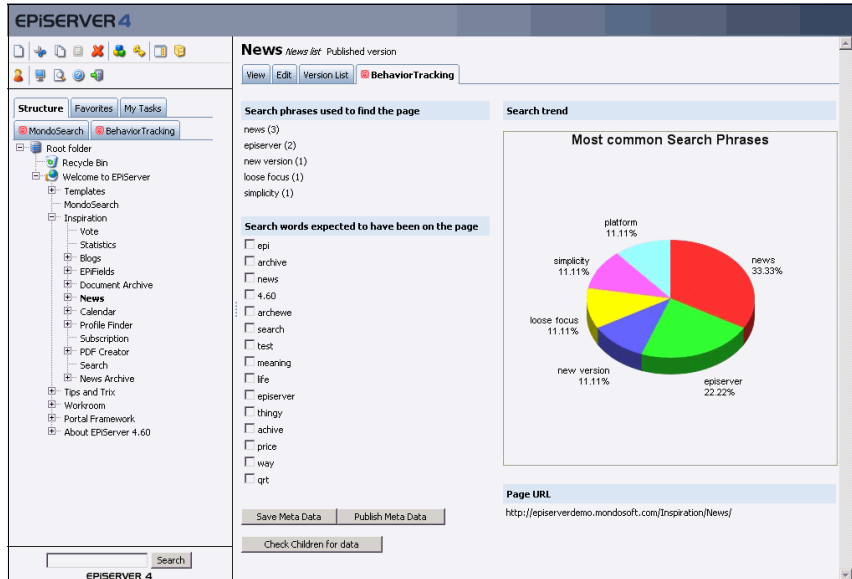
1. Log in to EPiServer as a user that has editing privileges.
2. Enter Edit Mode (e.g., by using the right-click context menu).
3. Use the standard EPiServer **Structure** tab in the left frame to locate and select the page for which you wish to see data.
4. Click on the **BehaviorTracking** tab in the right frame (in the editing area).

For more information about logging in to EPiServer and entering Edit Mode, please see your EPiServer documentation.

Note that you can also view the page details by clicking on a page link on one of the other views of BehaviorTracking (such as **Term Details**).

## Interpreting the Page Details

The single-page **BehaviorTracking** view shows information about which search words were used to find the current page and which additional words seem to have been expected by users who had to refine their query before finding the page. Controls are provided that enable you to add metadata that will associate one or more of the expected words with the current page.



### Search Phrases Used

The **Search phrases used to find the page** list identifies the terms most often used to find the current page, with the number of occurrences of each phrase shown in parenthesis. The **Search trend** display provides a graphical representation of this same information, giving a visual impression of how popular each search term was when finding the page.

Consider how the words listed here compare to the actual content of the current page. For example:

- If the top one or two words show a much higher count than the other words, then these indicate the most popular topic of the target page. If the page is long and/or contains information on a wide variety of topics, consider moving the less-popular topics to another page.
- If one or more of the words seem to indicate that the current page is not actually the page users wanted, then consider using keyword META-tags with these words on the better page.

### Reading and Adding Expected Terms for the Page

Under the **Search words expected to have been on the page** heading you can read a list of words that were often used in the same search session in which the current page was eventually selected. These indicate key words that users might have been expecting to find on this page.

To use this list, read each word and consider how it relates to the content of the current page. Then do one of the following:

- If the word is relevant for the current page and this page is a likely best match for that word, then mark the check box for that word to add it as a key-word META-tag for the page.
- If the word is a synonym or common misspelling, then consider making a synonym that matches the word to the correct spelling or the equivalent term more often used on your site. See the *InformationManager User's Guide* for details about how to define synonyms for MondoSearch.
- If the word is most relevant for another page, then add that word as a key-word META-tag on the better page using the standard tools of EPiServer (see your EPiServer documentation for instructions).
- If the word makes no sense to you, just ignore it.

After you are done inspecting the list, click on the **Save Meta Data** button if you have marked any of the check boxes (or click on **Publish Meta Data**, if available, to save and publish right away).

### Fetching and Interpreting Child-Page Data

If you would like to get an overview of which search words have led users to any of the pages in a selected branch of your site, then go to the BehaviorTracking tab for the page at the highest level of the branch you want to inspect and click on the **Check Children for Data** button. All of the relevant terms are then listed.

### Page URL

The **Page URL** value identifies the URL at which users will see the current page. This URL can otherwise be difficult to find while in Edit Mode.

## Acting on BehaviorTracking Data

This section provides some general advice for how to interpret and act upon the information provided by BehaviorTracking.

### When No Pages are Found

If you can see that a large portion (e.g., 20% or more) of searches are resulting in zero found pages, consider the following:

- Users may be confused by your search form. For example, they think the search submit button is actually a link to another page where they expect to find the actual search form.
- Users might misunderstand the purpose of your site. For example, you could have a site called *Network Doctor* for troubleshooting network problems, but users are looking for medical advice. Check the list of common search terms to see if this might be true (see “Top Search Terms” on page 93). If so, consider redesigning your home page to make the purpose of your site more clear. You might also check to see how your site is appearing in the larger, global Internet search engines.
- Your site might be missing the information that users are seeking. Consider adding information about topics that many users seem to be interested in. Check the list of top not-found words for ideas (see “Top Not-Found” on page 94).
- Users might not understand how to formulate a query. Check the list of search terms to see if this might be true. If so, consider including some extra instructions on your search-form page. See “Top Search Terms” on page 93.
- Users might often be misspelling their search terms. This might especially be the case when users search for products or other special words that they may have heard but are not accustomed to writing down. Check the list of top not-found words for common misspellings. Consider establishing common misspellings as synonyms to the correct word or adding them to the keyword META-tags for pages that are likely to help. See “Top Not-Found” on page 94.
- Your search engine might be interpreting user queries too literally. Try enabling phonetics, synonyms, stemming and auto-boolean logic to help queries return more pages.

### **When a Small Number of Pages is Found**

If a large percentage of search results include a small number of found pages (relative to the size of your site), then users might be formulating extremely accurate queries or the pages of your site might be wildly different, so that only a few pages match any given query. However, it might also indicate that the search engine is not returning enough matches to each query. If you can also see that there are a large number of searches with zero results, then try to find out why. Some possible reasons could be:

- The crawler might be having trouble grabbing the site, which means that the search database will not have all of your pages in it. Try checking the GrabMap of the MondoSearch InSite to see if this is the case, and, if so, make adjustments to your Crawler communication settings, starting points and/or login settings. Please see the *MondoSearch User's Guide* for details about how to work with the GrabMap.
- Your site might be missing the information that users are seeking. Consider adding information about topics that many users seem to be interested in. Check the list of top not-found words for ideas (see “Top Not-Found” on page 94).
- Users might be submitting odd search criteria—check the list of not-found words to see if this is the case (see “Top Not-Found” on page 94). Also consider if there is something about the content of your site that might be causing users to search in a

strange way; you may need to add extra advice to your search form page, or change the search options it gives.

- Your search engine might be interpreting user queries too literally. Try enabling phonetics, synonyms, stemming and auto-boolean logic to help queries return more pages. See “Configuring Search Options” on page 66 for details about how to enable these options.

### When a Large Number of Pages is Found

If a large percentage of search results include a very large number (e.g., more than 50) of found pages, users may tire of looking before they find what they are looking for. However, if MondoSearch is configured properly, then both the categorization feature and sorting algorithm should bring the most-wanted pages straight to the user’s attention.

Try to check the search-words lists and the results most often returned and followed from these words to see if you think the results are a likely match (see “Top Search Terms” on page 93). If you suspect that your users are getting swamped with too many search results, consider the following:

- It might help to improve your categorization. See Chapter 8: “Setting up MondoSearch Categories” for details.
- You might encourage users to enter more search terms into the form at once. Users may have had bad experiences doing this on other sites, while MondoSearch handles multi-word queries very well.
- You might need to add to the stopword list so common words for your site (e.g., your company name or other text that might appear in the footer of every page) will not trigger a hit. Please see the *MondoSearch User’s Guide* for details about stopwords and how to work with them.

### When Many Sessions Result in No Clicks

Try to compare the **Number of clicks** to the **Number of searches** on the **Overview - Searches per Day** page. If you notice that there are many more searches than clicks, then users seem to be leaving your search engine without finding what they were looking for. In this case, consider the following:

- New users might initially misunderstand the purpose of your site. When they see the types of pages returned, they realize this and leave rather than load a page
- Your page titles and/or descriptions might be inaccurate or misleading. Users might be finding pages, but they consider the page names and/or descriptions to be uninteresting even though the right documents were actually listed. See “Optimizing Page Titles and Descriptions” on page 42 for advice about how to improve these for your site.
- Users might be having trouble finding the right page in the result list—especially when result lists are long. In this case, you should improve your page-ranking system. Consider using (or improving) categories to make result lists easier to navigate (see Chapter 8: “Setting up MondoSearch Categories”). Also consider using rank-word META-tags and rank-document META-tags to help improve the way your best pages

are ranked. Please see the *MondoSearch User's Guide* for details about how to use META-tags to influence page rankings.

- Users might not be getting any results, so they have nothing to click on. Check the value for **Percent of searches with no results**; if this seems high, follow the advice given in “When No Pages are Found” on page 108.







# CHAPTER 12

## InformationManager Features in EPiServer



InformationManager actively connects your knowledge of your site's statistics and trends—obtained, for example via BehaviorTracking—to the way visitors are met in your online environment. It provides many advanced tools for tuning your search engine and applying the lessons you have learned by analyzing your site's log with BehaviorTracking.

The integration packages adds two of the most important features from InformationManager (SearchHeaders and SearchNames) to the EPiServer Edit Mode interface, making it easy for site editors and managers to work with these features from within the familiar EPiServer environment. This chapter explains how to work with SearchHeaders and SearchNames in EPiServer.

Other InformationManager features, including Synonyms, BehaviorMatch, filter translations and local-via-global search, can also be used with your EPiServer site, but these must be managed using the native InformationManager interface; please see the *InformationManager User's Guide* for details.

## Working with SearchHeaders

The SearchHeaders feature from InformationManager enables you to create banner displays that are triggered in response to specific queries. Use them to guide user interests, run campaigns, promote new products and to help make your search engine easier to use. See also “Targeted Promotions with SearchHeaders” on page 19 for an introductory discussion of SearchHeaders.

### Creating SearchHeaders

To define a query and specify a SearchHeader for it:

1. Go to Edit Mode in EPiServer, click on the **InformationManager** tab in the left frame and then click on the **SearchHeaders** link. This page lists all of the currently defined SearchHeaders and enables you to modify the list.

**EPISERVER 4**

Structure Favorites My Tasks

MondoSearch BehaviorTracing

**InformationManager**

SearchHeaders

SearchNames

English (US)

Search

**SearchHeaders**

SearchHeaders are special displays that are triggered by specific search words. They enable you to promote new products, special offers, company news and more based on stated user interests (i.e., their queries). SearchHeaders often take the form of banner ads placed at the top of the result list, but you can design them however you like. They can contain all types of HTML, so you can include text, graphics, animations and/or links as required.

Query	Exact	Personalize No.
BehaviorTracking	False	1
BehaviorTracing	False	2
InformationManager	False	1
InformationManager	False	2
EPiServer	False	1

Add SearchHeader

- Click on the **Add SearchHeader** button at the bottom of the table. The SearchHeader input form opens.

EPISERVER 4

SearchHeaders

Update the SearchHeader and save it. You can also delete the SearchHeader by clicking on the Delete button.

Item	Value
Query	<input type="text"/>
Exact	<input type="checkbox"/>
Personalize No.	<input type="text" value="1"/> Use "1" to create the default header.
Header (HTML)	<div> <div> B / U A [List Icons] [Table Icon] [Link Icon] [Unlink Icon] [Image Icon] [Media Icon] [Help Icon] </div> <div> Dialog 9 8 Heading 1 </div> <div> </div> </div>

Save Delete Cancel

English (US) Search

EPISERVER 4

- Make the following settings:
  - Query:** the word or phrase as you expect it to be submitted by a user. If more than one word is specified, then it will be interpreted as a phrase, which means that the query must include exactly those words in that order. You cannot define multiple queries in a single entry.
  - Exact:** if set to “true”, then the SearchHeader will take effect *only* when the query matches your exact text in the **Query** field, with no other words. When cleared, the SearchHeader will take effect for all queries that include the current entry (even if there are even more words in the original query, which themselves might also trigger additional SearchHeaders).
  - Personalize No.:** SearchHeader personalization enables you to create several SearchHeaders for the same query, leaving the selection of which SearchHeader to use until run time. If you are using personalization on your site, then you may have several SearchHeaders configured with the same query, but with different personalization numbers. If you are not using personalization, the personalization number must always be “1”. See also “Dynamic SearchHeaders” on page 116 for more information about this feature.
  - Header (HTML):** Use the editing tools provided here to create the content of the SearchHeader. The tools offered here are the same as those offered elsewhere for creating HTML in EPiServer; see your EPiServer documentation for full details.
- Click on the **Save** link to add your new entry to the list. (Click on **Cancel** or **Delete** to abandon your new entry without adding to the list.)
- Repeat this procedure until you have configured all of the SearchHeader entries you wish to use at this time.

## Editing SearchHeaders

To edit an existing SearchHeader, simply click on its name in the **Query** column. The selected entry then opens in the edit form. Here, you can edit the settings as required (see “Creating SearchHeaders” on page 114 for more information about the settings). Note, however, that you cannot edit the query for existing SearchHeaders; if you must edit a query, then you must delete the existing SearchHeader and create a new one with the new query word.

After making your changes, do one of the following:

- Click on **Update** to apply them to the displayed list (as usual, they are not saved on the server until you click on **Save**).
- Click on **Cancel** to revert to the settings that were in effect just before you clicked on Edit for the current entry.

## Deleting SearchHeaders

To delete an existing SearchHeader, click on its name in the **Query** column. The selected entry then opens in the edit form. Click on the **Delete** button here to remove the SearchHeader.

## Dynamic SearchHeaders

In most cases, SearchHeaders are “static” in the sense that they always appear the same each time the given trigger word is used in a search. However, it is possible to create several different SearchHeaders for a given trigger word and leave the decision of which to use until run time. In this case, your result page must be programmed to select the appropriate one based on whatever criteria you choose (e.g., preferred language, user access level, time of year, etc.). The various versions of a given SearchHeader are distinguished by their “personalization number”, which must be unique for all SearchHeaders that use the same trigger word.

Dynamic SearchHeaders are an advanced feature that requires custom .NET programming of your search/result page. If they are in use at your site, then please speak with your developers and/or system administrator to find out the details of your local implementation so that you can author your SearchHeaders correctly. If you are a developer and would like to make use of this feature, please see your MondoSearch developer documentation for details.

## SearchNames

The InformationManager SearchNames feature reacts to specific queries to bring users directly from the search form to the best matching content page, bypassing the result list entirely. Use the SearchNames interface in EPiServer to establish the relevant queries and the target URLs that they should trigger.

## Creating SearchNames

To define a query and specify a target URL for it:

1. Go to Edit Mode in EPiServer, click on the **InformationManager** tab in the left frame and then click on the **SearchNames** link. This page lists all of the currently defined SearchNames and enables you to modify the list.

Query	Exact	Personalize No.
Mondosoft	True	1

Add SearchName

2. Click on the **Add SearchName** button at the bottom of the table. The SearchName input form opens.

Item	Value
Query	
Exact	<input type="checkbox"/>
Personalize No.	1 Use "1" to create the default name.
URL to redirect to	

Save Delete Cancel

3. Make the following settings:
  - **Query:** the word or phrase as you expect it to be submitted by a user. If more than one word is specified, then it will be interpreted as a phrase, which means that the query must include exactly those words in that order. You cannot define multiple queries in a single entry.
  - **Exact:** if set to “true”, then the SearchName will take effect *only* when the query matches your exact text in the **Query** field, with no other words. When cleared, the SearchName will take effect for all queries that include the current entry.
  - **Personalize No.:** SearchName personalization enables you to create several SearchNames for the same query, leaving the selection of which SearchName to use until run time. If you are using personalization on your site, then you may

have several SearchNames configured with the same query, but with different personalization numbers. If you are not using personalization, the personalization number must always be “1”. See also “Dynamic SearchNames” on page 118 for more information about this feature.

- **URL to redirect to:** Enter the URL of the page that should load (instead of the result list) whenever a user searches for the trigger query.
4. Click on the **Save** link to add your new entry to the list. (Click on **Cancel** or **Delete** to abandon your new entry without adding to the list.)
  5. Repeat this procedure until you have configured all of the SearchName entries you wish to use at this time.

## Editing SearchNames

To edit an existing SearchName, simply click on its name in the **Query** column. The selected entry then opens in the edit form. Here, you can edit the settings as required (see “Creating SearchHeaders” on page 114 for more information about the settings). Note, however, that you cannot edit the query for existing SearchNames; if you must edit a query, then you must delete the existing SearchName and create a new one with the new query word.

After making your changes, do one of the following:

- Click on **Update** to apply them to the displayed list (as usual, they are not saved on the server until you click on **Save**).
- Click on **Cancel** to revert to the settings that were in effect just before you clicked on Edit for the current entry.

## Deleting SearchNames

To delete an existing SearchName, click on its name in the **Query** column. The selected entry then opens in the edit form. Click on the **Delete** button here to remove the SearchName.

## Dynamic SearchNames

In most cases, SearchNames are “static” in the sense that are always the same each time the given trigger word is used in a search. However, it is possible to create several different SearchNames for a given trigger word and leave the decision of which to use until run time. In this case, your result page must be programmed to select the appropriate one based on whatever criteria you choose (e.g., preferred language, user access level, time of year, etc.). The various versions of a given SearchName are distinguished by their “personalization number”, which must be unique for all SearchNames that use the same trigger word.

Dynamic SearchNames are an advanced feature that requires custom .NET programming of your search/result page. If they are in use at your site, then please speak with your developers and/or system administrator to find out the details of your local implementation so

that you can author your SearchNames correctly. If you are a developer and would like to make use of this feature, please see your MondoSearch developer documentation for details.







# APPENDIX A

## Using the Native MondoSearch Interfaces



The MondoSearch for EPiServer package adds many of the most-used configuration, information and management controls to your standard EPiServer environment. However, even more features and controls are provided by the traditional interfaces to each of the MondoSearch applications (MondoSearch, BehaviorTracking and InformationManager).

Though you will probably stick with the EPiServer interface most of the time, you might occasionally want to take advantage of the additional features provided through native interfaces to the various MondoSearch applications. Some of the most important extra features offered by these interfaces are outlined in the sections below. Settings that are common to both interfaces will produce the same effect regardless of which interface you use.

## Search Configuration in the MondoSearch InSite

The configuration and control interface for the MondoSearch search engine is called the *InSite*. It provides many tools for configuring, monitoring and controlling the Crawler and many settings that control how results are displayed and other features. The InSite remains fully functional after integration with EPiServer, so you can still use it at any time. All you need is the correct URL and an InSite user account. Talk to your system administrator and/or see the *MondoSearch User's Guide* for details.

Extra features and settings provided by the InSite include:

- *Detailed Crawler control*  
Many types of advanced settings help make sure the Crawler can get to all of the relevant areas of your site. Set multiple starting points, passwords for accessing restricted areas, form-follow options and more.
- *Fine-tune response for difficult sites*  
MondoSearch features detailed duplicate-document detection, correct indexing of documents that use HTML frames, and support for many specialized site-coding techniques. All of these features can be fine-tuned to ensure the system provides the best possible response for your specific site.
- *Detailed Crawler analysis with the GrabMap*  
The MondoSearch GrabMap enables you to view your site exactly as the Crawler saw it. Use the GrabMap to set new start points, mark “do-not-crawl” areas and for detailed troubleshooting.
- *Fine-tune the response to binary document formats*  
The InSite gives extra control over the way the Crawler indexes PDF, Office and Flash documents.
- *Detailed META-tag configuration*  
Configure the Crawler to recognize and handle the various types of META-tags in use on your site.

Note that many of the settings provided by the InSite affect the construction of CGI-based search and result pages generated by the system; these are not relevant when you are using EPiServer because in this case, your search and result pages are custom-coded .NET pages managed by EPiServer, rather than the CGI pages generated by MondoSearch. In general, you should simply ignore the entries under the **User Interface** menu of the InSite.

## Using the Stand-Alone BehaviorTracking Application

In addition to the BehaviorTracking tools provided in EPiServer (as described in Chapter 11: “Working with BehaviorTracking in EPiServer”), the stand-alone BehaviorTracking application provides even more tools and information, including:

- *Regular e-mail reports in Microsoft Excel format*  
BehaviorTracking can automatically deliver reports formatted as Microsoft Excel documents at regular intervals.
- *Expectation map*  
This graph that indicates how users may *expect* your site to work.
- *Date-span compare*  
Most of the views in BehaviorTracking are able to show data for two different date spans, enabling you to see how your data changes with time.
- *Watch groups*  
These enable you to monitor traffic from specific hosts, such as competitors, regulators or customers.
- *Application configuration*  
BehaviorTracking must be configured with the location of the log file generated by MondoSearch. This enables the BehaviorTracking import program to load the log into its database at regular intervals. This setting must be made after you install the integration package (see also “Required Initial Configuration Settings” on page 24) but usually will not change unless you modify your Web server installation.

The native BehaviorTracking interface remains fully functional after integration with EPiServer, so you can still use it at any time. All you need is the correct URL and a BehaviorTracking user account. Talk to your system administrator and/or see the *BehaviorTracking User’s Guide* for details.

## Using the Stand-Alone InformationManager Application

In addition to the InformationManager tools provided in EPiServer (as described in Chapter 12: “InformationManager Features in EPiServer”), the stand-alone InformationManager application provides even more tools and information, including:

- *Synonyms*  
Improve search results by matching common search words to local site equivalents. Use this to correct for common misspellings, map competitors’ trade names to your own equivalent products and to compensate for differences in terminology.
- *Online BehaviorMatch approval*  
Enables you to create new synonyms simply by marking check boxes in the list of suggested matches (see also “BehaviorMatch” on page 102).
- *Local-via-global search*  
Users arriving from a global search engine (such as Google) will automatically see both the page found by the global engine plus a list of other local pages found by the same query recreated in MondoSearch.

- *META-tag filters and translations*  
Create custom, field-based search forms in which users can search for specific types of information, such as author, subject, department, product type, etc. This information is found and classified based on META-tag data stored for each page. You can also configure META-tag translations to compensate for differences in the way various authors use META-tags, creating a system that is more compact, efficient and easy to search.

The native InformationManager interface remains fully functional after integration with EPiServer, so you can still use it at any time. All you need is the correct URL and a InformationManager user account. Talk to your system administrator and/or see the *InformationManager User's Guide* for details.



# APPENDIX B

## Advanced Integration and Development



This integration package delivers out-of-the-box functionality that provides a full-featured search and analytics package with basic customization features that do not require any programming. However, because the solution is based on standard .NET Web services and programming techniques, developers are also able create fully customized solutions by modifying the supplied code or even by developing their own features from scratch. This appendix gives a few ideas for advanced development projects and some guidelines for getting started with them.

## Setting Custom Search Options with MQL

All of the queries and commands submitted to MondoSearch from EPiServer are written using the Mondosoft Query Language (MQL). Developers designing their own pages using Microsoft Visual Studio must therefore program their pages to submit all requests to MondoSearch, BehaviorTracking or InformationManager as MQL messages.

In most cases, the MQL is hidden when you work with the integration features added to EPiServer. However, the standard search/result page provided with the integration package enables editors to customize search options by entering MQL commands directly into the **Edit** view of any instance of the MondoSearch page type (see also Chapter 7: “Search/Result Page Configuration in EPiServer”). For example, your site could include several different search pages, each targeted for a different purpose with behind-the-scenes MQL customizations.

MQL statements are human-readable and very easy to write. For example the following statement uses the standard “createdby” META-tag added by the Sitecore through the MondoSearch META-tag rendering to search for documents written by a specific author:

---

```
FILTERS CREATEDBY CONTAINS 'Allan'
```

---

A couple of other examples of using MQL are also given in “Optimizing Page Titles and Descriptions” on page 42.

Complete details of how to work with MQL for the MondoSearch SearchService are given in the *MondoSearch Developer’s Guide*. Please refer to that manual both for a general discussion of MQL and for the specific details about how to formulate MQL commands for the SearchService.

## Building Your Own Search/Result Pages

Because the MondoSearch integration is built entirely using standard .NET programming techniques and Web service communication, developers familiar with .NET can easily design their own search/result pages. The only requirements are:

- All requests must be sent as valid MQL to the MondoSearch SearchService
- The page must be prepared to handle the result set as it is returned by the SearchService.

The MondoSearch page type delivered with the integration package is simply a standard .NET page that generates both the search form and displays a result list. Developers can use this as a starting point for customization or simply as an example for beginning their own projects.

For example, it is possible to create a much more detailed “advanced search” form, in which users can select options for language, category, boolean logic, media types, various META-tag values and more. All you need to do is design an input control for each type of option and program it to generate the appropriate MQL statement.

Simple, cosmetic customizations to the supplied search/result page (such as changes in colors, fonts and graphics) can easily be made by editing a copy of that page in Microsoft Visual Studio.

Complete details of how to work with MQL for the MondoSearch SearchService are given in the *MondoSearch Developer's Guide*. Please refer to that manual both for a general discussion of MQL and for the specific details about how to formulate MQL commands for the SearchService.

## Custom Integration and Feature Development

In addition to working with the traditional search/result interface delivered with the integration package and discussed above, developers are also able to invent new ways of integrating search results and BehaviorTracking data into entirely new types of features.

One example of a new, unique feature that uses search data in a non-search-form environment is the *most-wanted results* feature provided with the integration package. This feature is based on the theory that the pages most-often selected from search results will probably be of interest to a large portion of all site visitors. Therefore, the home page of the site is enhanced to query BehaviorTracking and to display links to these documents right on the front page.

As with the search/result page, all that developers need to do is to design standard .NET pages that interact as needed with the various MondoSearch services as follows:

- To set search options and submit queries to the search database, use the MondoSearch Search Service (see the *MondoSearch Developer's Guide* for details)
- To issue commands to and read the status of the Crawler, use the MondoSearch Management Service (see the *Mondosoft Integration Services Developer's Guide* for details)
- To submit queries to the BehaviorTracking database, use the BehaviorTracking service (see the *Mondosoft Integration Services Developer's Guide* for details)
- To read and set synonyms, SearchNames and SearchHeaders, use the Information-Manager service (see the *Mondosoft Integration Services Developer's Guide* for details)

Note, however that many of these advanced customization options require that you license the full Mondosoft Integration Services package from Mondosoft (a limited, undocumented version is included to provide the default functionality in EPiServer). Please contact your Mondosoft representative for details.







# APPENDIX C

## Installing Mondosoft Integration Services



Mondosoft Integration Services is an add-on product for MondoSearch. It makes crawler-control, BehaviorTracking and InformationManager features available via Web services. Normally, this product is sold separately, but a limited version is included with the MondoSearch for EPiServer integration package. This limited version does not include documentation, but does provide all of the functionality required for the default integration scenario as described in this manual. If you do not have the full version, then please refer to this appendix for details about how to install the version included with this package.

## Requirements

Before you begin to install the Integration Services package, you must already have the following installed on your server:

- Functioning MondoSearch 5.2
- Functioning InformationManager 5.2
- Functioning BehaviorTracking 5.1b
- .NET Framework 1.1 and ASP.NET on the same server

Note that Integration Services 1.0 does not support MJ-licenses for the MondoSearch and InformationManager web services. You can use the BehaviorTracking web service, however.

## Package Contents

The Integration Services package is an add-on for MondoSearch, InformationManager and BehaviorTracking. The unzipped package contains:

- \install.txt
- \ConnectionBuilder\ConnectionBuilder.exe
- \ConnectionBuilder\ConnectionBuilder.dll
- \gac\Mondosoft.MSProvider.dll
- \gac\Mondosoft.IMProvider.dll
- \gac\Mondosoft.BTProvider.dll
- \WebService\MSService.asmx
- \WebService\IMService.asmx
- \WebService\BTService.asmx
- \WebService\Web.config

You must add several of these files to your current project or installation by following the instructions below.

## Installing the Files

1. Unzip the package to a temporary folder by running `is10cd.exe`.
2. Do one of the following:
  - Install the .NET assemblies in the Global Assembly Cache (GAC).  
Use the GAC only if you plan to access the providers via their web service front-ends. To do so, locate the 3 assemblies in the “gac” folder of the unzipped Integration Services package and add them to the GAC. You can do this *either* by dropping the assemblies into the global assembly cache folder (%WINDOWS%\Assembly) *or* by using the GAC tool with the /i option; for example:  
`gacutil /i Mondosoft.MSPProvider.dll`  
`gacutil /i Mondosoft.IMProvider.dll`  
`gacutil /i Mondosoft.BTProvider.dll`  
The “gacutil” tool is a part of the .NET Framework SDK.
  - Install the .NET assemblies without the GAC.  
If you do not wish to use the GAC, you can instead copy the assemblies located in the “gac” folder of the unzipped Integration Services package to the “bin” folder of the web application where you will use them. Alternatively, if you are developing a non-web application (i.e., you will be accessing the provider DLLs locally), then copy the assemblies to the folder where the other executables for your project reside.
3. If you will be using the providers via the web services, then you must place each of the ASMX files in a folder that is accessible through your IIS. Set up the ASMX files as follows:
  - Create a folder in your `wwwroot` with the name you want your web service to have (e.g., “IS”). Copy the ASMX and Web.config files from the “WebServices” folder of the unzipped Integration Services package to this new folder. If you are using Windows Server 2003, then verify that both the IUSR and the NETWORK SERVICE user have read and execute access to this folder and its sub-folders. Alternatively, if you are using Windows 2000 Server, then verify that both the IUSR and the ASPNET user have read and execute access to this folder and its sub-folders.
  - In the IIS console, choose **Properties** for the newly created folder and select **Create Application**.
  - For the BehaviorTracking WebService you should also edit the Connection String property in Web.config to match your SQL Server settings (i.e., insert the proper value in `<add key="Mondosoft.BehaviorTracking.DBConnectionString" value="" />` in Web.config). See also “Specify the Database Connection String for BehaviorTracking” on page 132 for complete details.
  - You should now be able to call each service from your browser by entering URLs such as:  
`http://[your-server-name]/IS/BTService.asmx`

4. Confirm that the web-services user has access to the Registry Database (for Windows Server 2003, this user is typically called “NETWORK SERVICE”; for Windows 2000 Server, “ASPNET”).

The web-services user must have write access to the following registry key:

---

HKEY\_LOCAL\_MACHINE\Software\Mondosoft

---

To check the access control list (ACL) for this key, do one of the following:

- For Windows Server 2003:  
Run **regedit.exe**, navigate to HKEY\_LOCAL\_MACHINE\Software\Mondosoft, right-click on the **Mondosoft** key and select **Permissions**. The familiar Windows ACL Permissions dialog box is now shown. From here, you can set or change the access permissions of the **NETWORK SERVICE** user.
- For Windows 2000 Server:  
Run **regedt32.exe**, navigate to HKEY\_LOCAL\_MACHINE\Software\Mondosoft, open the **Security** menu and select **Permissions**. The familiar Windows ACL Permissions dialog box is now shown. From here, you can set or change the access permissions of the **ASPNET** user.

## Configuring the Installation for Use

### Specify the Database Connection String for BehaviorTracking

When you use the BehaviorTracking Provider, you must configure your application or web service with a connection string for the BehaviorTracking database. The required string takes the following form:

---

User ID=<user>;Password=<pass>;Data Source=<server>;Initial Catalog=<database>;

---

Where:

- <user> is the user name of an SQL server account that has full access rights to the BehaviorTracking database.
- <pass> is the password for the account named above.
- <server> is the machine name of the computer where the database is running
- <database> is the name of the database itself running on the above-named server. Usually this is “BehaviorTracking”

For example:

---

User ID=bt;Password=bt;Data Source=cphdev00;Initial Catalog=BehaviorTracking;

---

Depending on your specific network and/or SQL Server installation, you might need to modify or extend the connection string slightly according to the standards for SQL Server

(see your SQL Server documentation for details). The required string is the same as the one that should already be stored in the **Web.config** file for your BehaviorTracking installation.

To use the BehaviorTracking web service, enter the database connection string into the **Web.config** file for the service as follows:

1. Open the folder on your web server to which you copied the BehaviorTracking service and its initial **Web.config** file.
2. Open the **Web.config** file in a text editor and find the following line:

---

```
<add key="Mondosoft.BehaviorTracking.DBCConnectionString" value=""/>
```

---

3. Enter your connection string as the value of this key.

To use the BehaviorTracking Provider from a local application rather than via the service, enter the database connection string into the configuration file for your application (usually, you will keep this in the same folder as the application itself or in the same folder as the BehaviorTracking Provider DLL). Do the following:

1. Open your application configuration file in a text editor.
2. Add the following line:

---

```
<add key="DBCConnectionString" value="" />
```

---

3. Add your database connection string as the value of the new key.

Note that the database connection string described above is not the same as the provider connection strings described in “Connecting to the Providers” on page 16. Both types of strings are required.

## Create an Information Manager User if Needed

The InformationManager web service requires that at least one InformationManager user is created. We recommend that you use the same login and password as the MondoSearch user.

Please see the InformationManager documentation (near the end of the “Installing InformationManager” chapter) for details about how to create and manage users in InformationManager.



# Contacting Mondosoft

## Europe

Mondosoft A/S  
Vestergade 18E  
DK-1456 Copenhagen K  
Denmark

Phone: +45 70 20 20 09  
Fax: +45 70 20 25 09

## USA

Mondosoft Inc.  
425 Sherman Avenue, Suite 300  
Palo Alto, CA 94306  
USA

Toll free: +1 (800) 625 1175  
Phone: +1 (650) 462 2140  
Fax: +1 (650) 462 2150

## Internet

Web: [www.mondosoft.com](http://www.mondosoft.com)  
E-mail: [info@mondosoft.com](mailto:info@mondosoft.com)

## Europe

Mondosoft A/S  
Vestergade 18E  
DK-1456 Copenhagen K  
Denmark

Phone: +45 70 20 20 09  
Fax: +45 70 20 25 09

## USA

Mondosoft Inc.  
425 Sherman Avenue, Suite 300  
Palo Alto, CA 94306  
USA

Toll free: +1 (800) 625 1175  
Phone: +1 (650) 462 2140  
Fax: +1 (650) 462 2150

## Internet

Web: [www.mondosoft.com](http://www.mondosoft.com)  
E-mail: [info@mondosoft.com](mailto:info@mondosoft.com)

