

CrownPeak Playbook▲

CrownPeak Search

Version 0.94

Table of Contents

- Search Overview 4**
- Search Benefits 4**
- Additional features 5**
- Business Process guides for Search Configuration 5**
 - Search Limitations..... 6
- Configuring Search 7**
 - Crawl the Site. 8
 - User Acceptance Testing (UAT)..... 8
- How Do I Use Search? 8**
- Making changes to Search implementations 9**
- Implementation Guidelines..... 9**
- Notes and Additional Information..... 9**

© 2014 CrownPeak Technology, Inc. All rights reserved. No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from CrownPeak Technology.

Document History

Author/Editor	Date	Reason for Change	Version
Corey Casado	12/20/2013	Draft	0.90
Fahd Shaaban	01/02/2014	Edits and Comments	0.91
Corey Casado	01/03/2014	Updates and new content based on initial review	0.92
Corey Casado	02/18/2014	Update and new content based on playbook meeting and additional feedback	0.93
Corey Casado	2/28/14	Update to search configuration guidelines	0.94

Search Overview

CrownPeak subscribers that have any of the following needs should consider implementing Search:

- Provide users with the ability to search for content on their site.
- Generate XML / RSS / JSON feeds based on keywords.
- Return dynamic / related content based on tags / taxonomy.

Search provides features that enable content teams to effectively leverage their investment in CrownPeak to meet these challenges.

Search Benefits

Search functionality within the site

- The ability to structure the search results independently of their content. This is useful when you want a page or an asset to appear in search results and the search terms are not contained anywhere within the content of the page or asset. This is accomplished through metadata fields and search tags. (See “Search Tag Reference” in Additional Information section)
- The ability to handle related and categorized content. When a page or asset is recognized as a search result the search can be configured to display any related content in the search results as well. This relationship is also managed through metadata fields and search tags. (See “Search Tag Reference” in Additional Information section)
- The ability to hide pages from search results. Search can be configured to not index certain pages and not display those pages in search results
- The ability to control which parts of pages are indexed. In situations where a keyword might be contained on every page in a Website such as *Home* or *Contact us* via a navigation wrapper or master page Search has the ability to ignore these sections during indexing so that search results are not filled with irrelevant pages.
- The ability to limit search results based on user types. This is useful for situations where the customer has pages or sections of their Website that should only be viewable by certain groups. For example, if a site has two types of users: business users and engineering users, and you don't want any engineering specific pages to appear when business users search for something. You can configure search to return business specific pages even when engineering pages contain relevant data as a result of the search query.

Search functionality on external sites

- Search functionality on external Websites known as *Federated search* is the ability to have search results that come from external Websites. As long as we are able to access the URL through *http* or *https* search can index the content of that Website.

Note: We will not have as much flexibility in terms of modifying metadata or other pieces of information on the external Website unless we are given access to the back-end or are able to communicate with someone who manages the external website.

Generation of XML / RSS / JSON feeds

- Templates within the Search UI allow us to customize how search results are rendered. Using these templates you can output the results in any format you need. Most often this is used to render XML, RSS, or JSON feeds but any text format is possible.

Additional features

CrownPeak Search has the following additional features.

- The ability to index the content of certain binary files. Search can index the content of the following binary files: .txt, .rtf, .xls, .ppt, .doc and PDF. If any of the search terms match the content of a binary file that file will be returned as a result.
- The type ahead feature that enables a user to enter text into the search textbox and a drop-down list displays with similar words and searches.
- The “Did you mean this?” feature suggests alternate spellings and search terms that closely match the original search terms.
- A Tag-based or non-user-entered related content search feature allows you to configure search so that results are only returned that match specific requirements such as particular sections of a page or specific metadata tags. This is used for a customized approach to search.
- Search promotions allow you to return a pre-defined chunk of HTML at the top of the results when a user enters a specific keyword or phrase. This is useful when you want to promote a particular product or new page on your Website.
- Synonyms allow you to associate specific words with one another; when a user searches one word, all other associated words are included in the search. For example, a user might want the words feline and kitty to be included in search results when a user enters the word cat.

Business Process guides for Search Configuration

Before starting the configuration, it is beneficial to solidify the scope of the configuration by collecting the following information:

- Verify that the URL starting points provided to the crawler contain anchor tags. This is typically a home page with links or an XML document that contains a list of anchor tags. This is important because Search will use these initial pages to find other pages to index. It is a good choice to provide a link to an XML document that contains links to several important pages so that it is not left up to the crawler to find these pages.
- Determine the type of authentication for the site. If the site requires an unsupported type of authentication then it will need to provide another method of access such as a separate HTTP authentication portal, user agent string, or cookie based authentication before we can crawl and index the site. Search directly supports HTTP and cookie based authentication

methods. Search does not directly support sites that use other types of authentication, such as Windows NTLM.

- Gather specific search requirements:
 - Is the search limited based on the type of user?
 - Are there different types of searches or different search result layouts? Currently an individual search instance can only implement 12 different layouts. If more are required then a ticket will need to be submitted to IT.
 - What sites should search be configured to crawl? Is Federated Search a requirement?
 - What languages need to be supported? Is double-byte or multi-byte search required? Search does not support double or multi byte pages. If this is a requirement then other alternatives will need to be explored.
 - Are there any filtering or sorting requirements?
- Verify that the site is complete and that there are pages, such as a site map, that allows the search engine to crawl effectively. *Search works by using a few initial URLs that redirect to HTML pages which contain links to other pages within the site. If there are pages that the customer wants to appear in the search results that are not “reachable” from the any of the initial URLs that are provided then those pages will not appear in the search results.*
- Would the customer like to use metadata? Such as PDF or Word documents.
- Verify that the site is published and that there are URLs available to crawl.

Search Limitations

It is also important to note some inherent limitations within search.

- Search is currently limited to displaying a maximum of 200 search results.
- Search is limited to 10 pages of search results.
- Search does not support double-byte or multi-byte search. If either of these are required other options should be explored.

Configuring Search

Once the Business Process is understood, use the following steps to configure Search:

1. Submit a ticket to IT to set up the instance(s)
 - Determine how many instances should exist for the site. It is strongly recommended that there is at least one instance per publishable stage in the workflow. This would mean that for most sites a stage and live instance are recommended.
 - Once the number of instances is determined, provide a search instance name for each instance. The name should also indicate if it is staging or a live site.
 - URLs to crawl for each instance.
2. Receive the HTML templates for the search results page. This step is important because it allows the developer to focus on the content of the search results rather than the look and feel.
3. Configure search instance to utilize metadata if applicable
 - To accomplish this for regular pages:
 - Setup the content fields to recognize the metadata (Configure > Textual > Content Fields).
 - To accomplish this for PDFs and other Binaries:
 - Each binary file will need to have an associated XML file published with the same name as the binary except for the extension. For example, *abc.pdf* will need to have a metadata file called *abc.xml*.
 - Under Configure > Crawler > URL MetaFile select **Replace Extension of existing URL** then in (1) MetaFile Extension add *xml*
 - In Configure > Textual > Content Fields, configure search to recognize the data to pull from the metafile. Be sure to use URL MetaFile in the Source drop-down.
 - For both Regular pages and PDFs after the content fields have been setup, re-crawl the site.
4. Configure applicable search templates. This is done in the search UI by clicking on **Templates** and then selecting and modifying the template that you want to use to render the output of search results. Often times only you only need to customize one search template. However if multiple searches are used, such as search results for products vs. search results for engineering documentation then multiple templates might need to be configured. For more information on how to customize search templates refer to the CrownPeak Search Guide on Connect.

Note: When configuring the template within the CMS that will query and display search results **do not configure the search URL to go to a specific search instance**. It should always point to <http://search.crownpeak.com/> and not something like <http://search.crownpeak10.com/>.

For example,

```
document.write('<'+script type="text/javascript"
src="http://search.crownpeak.com/cpt_search/result_1'+location.search+'&format=js&account=abcdefg
hijk"></'+script>');
```

Crawl the Site.

- Open the search UI and Click on Start Crawler => Start Crawling. There are several options on how to crawl.
 - **Cache Only** is the fastest option and will only crawl the cached files and will not check the Website to see if it has been modified.
 - **Cache** is the second fastest option and will use cache files only if no pages have been modified.
 - **Ignore Cache** will crawl the site from scratch. This is the slowest option but the most thorough as it does not rely on any caching mechanisms to optimize search indexing performance.

User Acceptance Testing (UAT)

- Testing occurs during the UAT portion of the development lifecycle because content needs to be published before search can crawl and index the site.
- Testing is accomplished by looking at all of the different configurations that were made and testing a few cases for each type of configuration. For example, the search was configured to use metadata fields, filter by user type, and generate XML feeds. Testing would comprise of searching for pages that should return as a result of the known metadata for those pages, searching using accounts with specific user types, and checking to make sure that the XML feeds are generated correctly.

How Do I Use Search?

Once the configuration is complete, CrownPeak will conduct user training as a part of User Acceptance Testing.

CrownPeak will review Search in the context of the business process, confirming the proper operations for:

- Searching, Filtering

CrownPeak will look at all of the user specific configurations that were implemented for this search instance and execute test cases that cover each configuration to verify that all the configurations are working as expected.
- Returning XML / RSS / JSON feeds

CrownPeak will look at the feeds that are generated by search and verify that those feeds match what is expected.

Any issues or additional changes requested will be processed as a part of the final configuration tuning.

Once the User Acceptance Testing is completed, normal operations will commence, the configuration project will end, and operations will be supported by CrownPeak's Absolute Service team.

Making changes to Search implementations

Currently within Search there is no programmatic association between instances. Unlike within the CMS where there is a workflow process that can be configured there is no workflow process in Search and every instance is completely separate from the other even if they are for the same CMS. Because of this the following process is **required** for *any* changes that are made to *any* Search instance.

To make a change to an existing Search instance:

1. The client must submit a ticket to support@crowpeak.com indicating the change to be made.

Changes to Search templates and or Search configurations should *first* be applied to the lowest publish state supported by the client. For example, Dev, Authoring, or Stage depending on the client.

2. Changes should be noted in the original ticket for future reference.

Once changes are applied client should be notified that the changes have been applied and to which state/url. Ticket should be marked as "Pending Confirmation". Ticket will close per normal ticket process.

3. The client should submit a new ticket requesting the changes be applied to the next appropriate state. Each state change should be accompanied by a new ticket.

The new ticket requirement can be waived if the client replies to the original ticket within 2 business days confirming the changes and requesting they be applied to the next state.

No changes should be applied to subsequent states without written approval by the client. However, for clients that open a ticket, request a change, and then say apply to live we don't necessarily need a second ticket that is the waive comment.

Implementation Guidelines

A Search project usually takes from 8 to 16 hours to configure. IT requires a 5 day lead time to create the instances. After the instances have been created configuration takes from 4 to 8 hours for the first instance with additional instances taking 2 to 4 hours. UAT depends on the complexity of the search implementation a complex implementation requires more testing.

Notes and Additional Information

More information can be found on Connect:

- CrownPeak Search guide: <http://connect.crownpeak.com/documents/developers/crownpeak-search-guide>
- Search Implementation Best Practices: <http://connect.crownpeak.com/documents/best-practices/searchimplementationbestpracticespdf>

- Search Tag Reference: <http://connect.crownpeak.com/documents/developers/search-tag-reference>