

Contents

Introduction	2
visionSearch Management Site Guide (IT)	2
User & Security Role.....	2
Search Engine.....	2
How to Setup a Search Engine	3
Collect Requirements.....	3
Create a search engine.....	3
Search Engine Indexing Activities.....	4
Integration with CMS6.....	7
Configure Vision Search (Advanced)	7
System Setting	7
System Variables.....	7
Content Type Rule (Advanced)	7
Security Role (Basic)	7
Synonyms (Basic)	8
Vision Search Widget (Basic).....	8
Note	8
Process Permission.....	8
IT.....	8
Advanced.....	8
Basic.....	8

Author

Yuliang Ding
Vision Technology Solutions, LLC.
June 1, 2015 – 2015.2.601.0

Introduction

visionSearch is built in-house on top of the popular open source search library Solr which is in turn built on Apache Lucene™. Solr is a very popular search platform used on such websites as eHarmony, Sears, Zappos, AT&T, Ticketmaster, eBay, Instagram, Netflix, Disney, Adobe and StubHub. Building our own search engine on top of Solr makes sense because we can rely on the powerful search technology built into Solr to ensure fast and accurate search results as well as intelligent indexing, while also being able to build and customize and continue to expand the functionality to work specifically with CMS6.

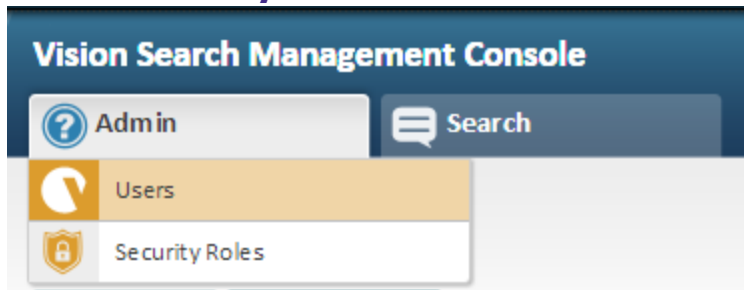
Vision Search includes two parts:

1. Search Engine Management Site
2. Integration with CMS6

visionSearch Management Site Guide (IT)

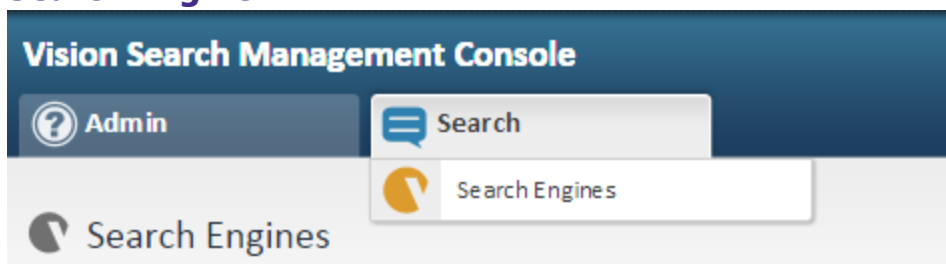
You can access visionSearch live site by: <http://search.visioninternet.com/>. Currently, only the IT team has the permission to access this site.

User & Security Role



The user and security role logic is the same as CMS6, if you want to add more users into Vision Search Management Site, please add related security roles and assign to the users.

Search Engine



We can create a search engine for each site, or for each different purpose of each site, e.g. different departments in one city may need different search engines. These search engines' settings won't be affected by each other.

How to Setup a Search Engine

Collect Requirements

Get the requirements from client:

1. What site URL does the client want to create search engine for?
2. If the client wants to only index specific pages, instead of whole website, please get the requirement that what kinds of pages on the site does the client want to index? E.g. all subpages of one page ...
3. About the indexing frequency, we will set it weekly by default. I don't suggest to ask client weekly or daily, as daily will use more CPU.

Based on above requirements, we can know:

1. Root URLs
Which URLs we should start to crawl the site.
2. Allow Paths
Which kinds of URL paths we can allow to crawl.
3. Indexing Allow Paths
Which kinds of URL paths we can allow to index.

Create a search engine

1. Go to Vision Search Management Site, Search -> Search Engines, Click "ADD SEARCH ENGINES" button:

The screenshot shows the 'Vision Search Management Console' interface. At the top, there are navigation tabs for 'Admin' and 'Search'. Below these, there is a 'Search Engines' section with a sub-tab also labeled 'Search Engines'. In this section, there are two buttons: 'DELETE SELECTED' and 'ADD SEARCH ENGINE'. The 'ADD SEARCH ENGINE' button is highlighted with a red box. Below the buttons is a table with the following data:

<input type="checkbox"/>	Name	Documents	Last Action Run
<input type="checkbox"/>	Gold CMS6_DEV	269	06/23/2015 12:22 PM
<input type="checkbox"/>	GOLD_CMS6_DEV	4325	03/08/2015 9:08 AM

At the bottom left of the screenshot, there is a small orange number '1'.

2. Put the basic information in the form:
 - a. **Name:** is required, and should be unique.

- b. **Description:** put any information here which is helpful to understand the purpose of the search engine.
- c. **Root URLs:** The URLs we should start to crawl the site, which is multiple lines; we can add multiple URLs, even they are in different sites.
- d. **Allow Paths:** The URLs paths or regular expressions we can allow to crawl, which is multiple lines; we can add URL path or regular expression for each line.
- e. **Disallow Paths:** The URLs paths or regular expressions we don't allow to crawl, which is multiple lines; we can add URL path or regular expression for each line. Usually, for CMS6 website, we don't need change this setting, just use the default one.
- f. **Indexing Allow Paths:** The URLs paths or regular expressions we can allow to index, which is multiple lines; we can add URL path or regular expression for each line.
This setting is different from "Allow Paths". Only when the indexed pages are different from the crawled pages, we have to setup this setting, otherwise, leave it as EMPTY.
- g. **Allow Formats:** The file formats we can allow to crawl and index.
For CMS6 website, usually, we don't need to change this setting, just use the default one.

After you have setup all the settings based on our requirements, click "SAVE" button. After you have saved, it will redirect back to the search engine list page.

Some settings will be generated by default, e.g. Search Engine Key, Search Engine Secret, Content Type Rule, Scanner Settings, Proxy Settings. Usually, we don't need to change these settings for our CMS6 websites, these settings have been optimized for our CMS6 websites.

Search Engine Indexing Activities

After we have created a search engine, we have to setup the scheduled search engine indexing activities.

1. Click "Edit", choose "Indexing" tab to setup the scheduled indexing activities.

Name	Documents	Last Action Run	Last Updated	By	Edit	Delete
Demo Search Engine	0		06/29/2015 8:15 AM	System, Admin		
Gold CMS6_DEV	269	06/23/2015 12:22 PM	06/23/2015 12:12 PM	System, Admin		
GOLD_CMS6_DEV	4925	08/08/2015 9:08 AM	02/27/2015 8:30 AM	System, Admin		

Displaying items 1 - 3 of 3

The screenshot shows the 'Search Engine Setting' page for a 'Demo Search Engine'. The 'Indexing' tab is selected and highlighted with a red box. Below the tabs, there are several form fields: 'Name *' (Demo Search Engine), 'Description' (This is the search engine for demo.), 'Language *' (English), 'Root URLs *' (dev5.visioninternet.com/CMS6_DEV/), and 'Allow Paths *' (*). Below these is the 'Schedule Setting' section, which includes dropdown menus for 'Index', 'Refresh', 'Optimize', and 'Clear', along with 'BACK' and 'SAVE' buttons. A red box highlights the 'Schedule Setting' section, and a callout bubble points to it with the text 'Frequent/One time Scheduled Indexing Activities'. Another red box highlights the 'Scheduled Activities' section, which contains 'DELETE SELECTED', 'SCHEDULE NOW', and 'REFRESH' buttons, with a callout bubble pointing to it and the text 'Schedule now'. Below the 'Scheduled Activities' section, it says 'No Search Engine Indexing found.'

Here, we have two options to setup the scheduled indexing activities.

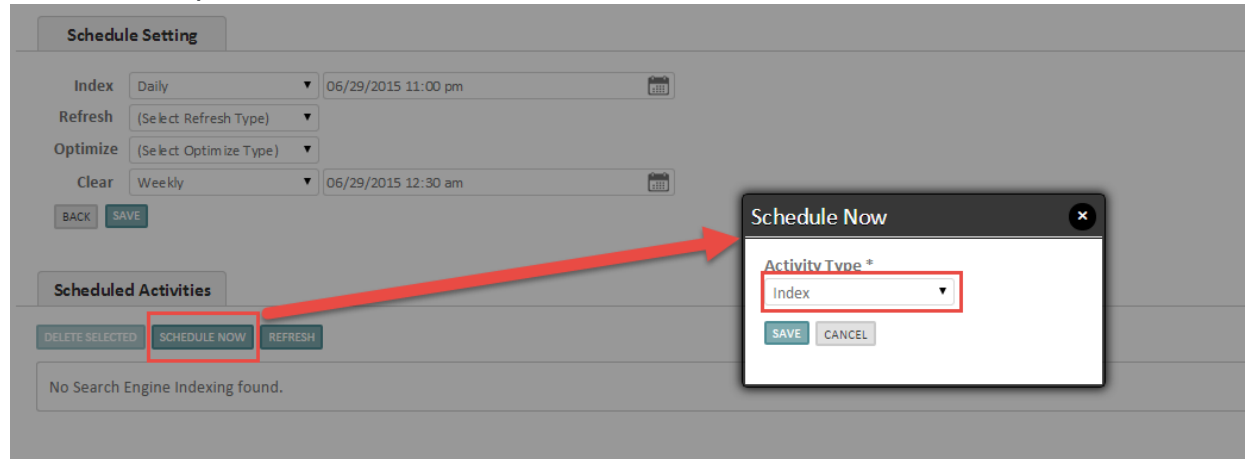
- a. Setup recurring/one time scheduled indexing activities.

This is for future scheduled indexing activities, we can setup a one-time schedule, or a recurring schedule. You have to click the "SAVE" button after you have chosen the scheduled options.

The screenshot shows the 'Schedule Setting' section. It has four rows of settings: 'Index' set to 'Daily' with a date/time picker for '06/29/2015 11:00 pm'; 'Refresh' set to '(Select Refresh Type)'; 'Optimize' set to '(Select Optimize Type)'; and 'Clear' set to 'Weekly' with a date/time picker for '06/29/2015 12:30 am'. At the bottom are 'BACK' and 'SAVE' buttons.

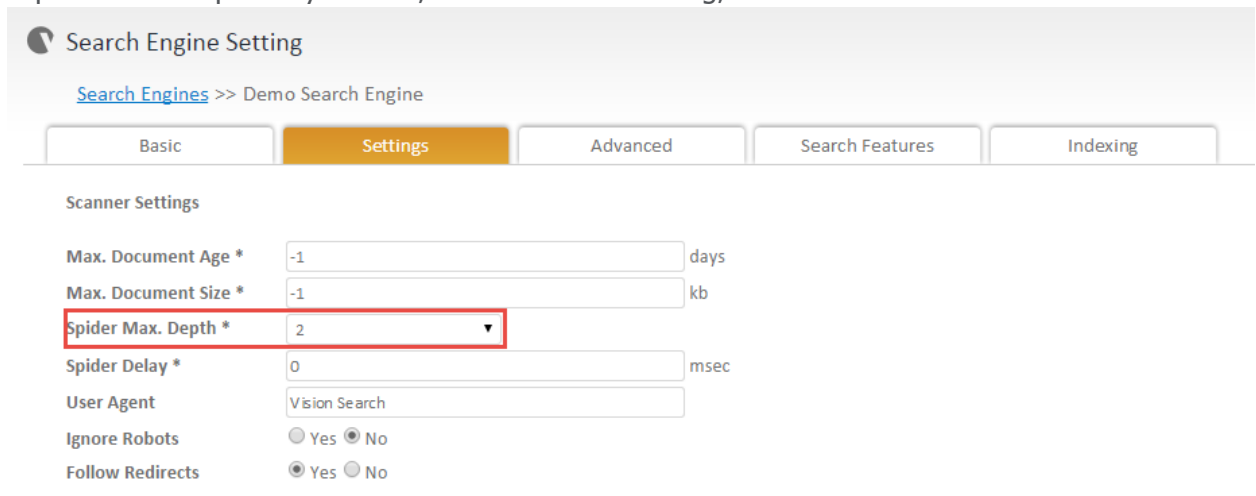
- b. Setup "schedule now" indexing activities.

After we have setup "schedule now" indexing activities, the search engine will run this activity ASAP.



CMS6 Best Practice:

- a. Once we have created the search engine, please "schedule now" at least one time.
 - b. For the Scheduled Indexing activities, please :
 - Index -> Daily -> the date with time after 11:00 pm.
 - Clear -> Weekly -> the date with time after 12:00 pm.
2. Now, we have finished the setup of a search engine. But if the site is big, which means there are lots of pages in that site, I would suggest to make another change to the search engine:
 "Spider Max. Depth": by default, it's 2. If the site is big, we can set it as 3 or 4.



Integration with CMS6

Configure Vision Search ([Advanced](#))

System Setting

Please enable the "Vision Search" system setting, if you want to use Vision Search in CMS6. By default, it's disabled.

System Variables

There are 3 system variables we have to setup for Vision Search, after "Vision Search" system setting has been enabled.

Search Engine Key	Sysadmin	System	[REDACTED]	05/17/2015 11:00 pm	System, Admin			
Search Engine Secret	Sysadmin	System	[REDACTED]	05/17/2015 11:00 pm	System, Admin			
Vision Search Server URL	Sysadmin	System	http://dev5.visioninternet.com/Search_DEV/	06/24/2015 11:47 pm	System, Admin			

1. Vision Search Server URL:
The URL of our Vision Search Management Site.
2. Search Engine Key:
The key of that search engine, please get from IT Team.
3. Search Engine Secret:
The secret of that search engine, please get from IT Team.

Content Type Rule ([Advanced](#))

Content Type Rule is used to categorize our frontend pages into different content types.

By default, once we have created the search engine, Vision Search Management Site will create the default Content Type Rule for CMS6 website, so, we don't need to change this.

But if we have any customized components, which its frontend component detail page URL format is totally our default one (~/Home/Component/{Component Name}/{Content Type}/{Content ID}/{Page ID}), please add 1 rule for the customized component.

You can sort the content type rule by drag & drop, Vision Search will try to match the rule from top to bottom, once it has found any rule matched, it will stop and use the matched one.

Content Type Rule ?

Search Synonyms		Content Type Rule			
Name	Content Type	URL	Edit	Delete	
Default		/Home/Components/{[a-z0-9+]}/?<contenttype>{[a-z0-9+]}/\{d+\}([a-z0-9]{32})/\{d+\}?			
Meeting Manager		/Home/Components/MeetingsManager/?<contenttype>[a-z0-9+].*			
Document	Document	/Home/ShowDocument.*			

Security Role ([Basic](#))

We have one security role action for Vision Search Management, for the regular user, who has the security role action will be able to setup the synonyms.

Synonyms ([Basic](#))

Please follow CMS6 Online Reference: Site Settings -> Site Configurations -> Search Management -> Content Type Rule.

Vision Search Widget ([Basic](#))

We have "Search Widget", which you can choose the Vision Search settings to create a Vision Search Widget. More details, please follow CMS6 Online Reference.

Note

Process Permission

IT

IT team is responsible for this process.

Advanced

SysAdmin (developer users) is responsible for this processes.

Basic

Regular CMS user is responsible for this processes.