



Yuliang Ding Vision Technology Solutions, LLC. Oct 5, 2015 – 2015.3.928.0

Table of Contents

Introduction	3
Vision Search Guide	3

Introduction

Vision Search 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

Vision Search Management Site Guide

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

User & Security Role

Visio	on Search Manage	ment (Console
0	Admin		earch
	Users		
0	Security Roles		

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?
- 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:

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

Create a search engine

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

visionLive		
Vision Search Manage	ement Console	
Admin	Search	
C Search Engines DELETE SELECTED ADD SERACH	ENGINE Documents	Last Action Run
Gold CMS6_DEV GOLD_CMS6_DEV	269 4325	06/23/2015 12:22 PM 03/08/2015 9:08 AM
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.

C	Search Engines						
DELET	E SELECTED ADD SERACH ENGINE						
	Name	Documents	Last Action Run	Last Updated	Ву	Edit	Delete
	Demo Search Engine	0		06/29/2015 8:15 AM	System, Admin	Ø	Θ
	Gold CMS6_DEV	269	06/23/2015 12:22 PM	06/23/2015 12:12 PM	System, Admin	Ø	•
	GOLD CMS6 DEV	4325	03/08/2015 9:08 AM	02/27/2015 8:30 AM	System, Admin	Ø	Θ
1						Displayi	ing items 1 - 3 of 3

Search Engine Setting

Basic Settings Advanced Search Features Indexing Name * Cemo Search Engine Description This is the search engine for demo. English Rott URLs * devis vision internet.com/CMS5_DEV/ Allow Paths * * changing >> Demo Search Engine Basic Settings Advanced Settings Advanced Settings Advanced Setting <						
Name * Demo Search Engine Description This is the search engine for demo. Language * English Construction Constr	Basic	Settings	Advanced	Search Features	Indexing	
Demo Search Engine This is the search engine for demo. Inguage * English Root URLs * dev5.visioninternet.com/CMS5_DEV/ Allow Paths * * * Allow Paths * * Setting Advanced Search Frequent/One time Scheduled Indoxing Activities Schedule now Schedule now Schedule now	Name *					
Description This is the search engine for demo. Language * English • Root URLS * dev5.visioninternet.com/CMS5_DEV/ Allow Paths * * Allow Paths * * Easis Settings Advanced Search Features Indexed Search Features Indexed Search Features Indexed Search Features Indexed Search Features	Demo Search Engine					
This is the search engine for demo. Language * English Rot URLs * dev5.vision internet.com/CMS6_DEV/ Allow Paths * * * * * * * * * * * * *	Description					
Language * English Root URLs * dev5.visioninternet.com/CM56_DEV/ Allow Paths * * * arch Engine Setting arch Engine Setting Advanced Search Features modern Basic Settings Advanced Search Features modern teach Seatting • Frequent/One time Scheduled Indexing Activities * * * * * * * * * * * * *	This is the search engine for dem	0.				
Language * English Root URLs * dev5.visioninternet.com/CMS5_DEV/ Allow Paths * * arch Engine Setting arch Engines >> Demo Search Engine Basic Settings Advanced Search Features Indexing checkuls Setting The UE det Index Type • Frequent/One time Scheduled Indexing Activities Frequent/One time Scheduled Indexing Activities Clear (Sect Clear Type) • © SCHEdule now eduled Activities				8		
Root URLs * dev5.visioninternet.com/CMS6_DEV/ Allow Paths * * arch Engine Setting Basic Settings Advanced Search Features Indexing CodeLe Setting Basic Settings Advanced Search Features Indexing CodeLe Setting The Gette Index Type • Frequent/One time Scheduled Indexing Activities Schedule now eciled Activities Schedule now	Language *					
Root URLs * dev5.visioninternet.com/CM56_DEV/ Allow Paths * * arch Engine Setting arch Engines >> Demo Search Engine Basic Settings Advanced Search Features Indexing thedule Setting thedule Setting Frequent/One time Scheduled Indexing Activities imite Elect optime Type • imite Elect opti	English .					
Allow Paths * Allow Paths * * Allow Paths * *	Root URLs *	2511				
Allow Paths * * * * * * * * * * * * *	dev5.visioninternet.com/Civis6_i	DEV/				
	Allow Paths *			6		
arch Engine Setting arch Engines >> Demo Search Engine Basic Settings Advanced Search Features Indexing checkule Setting Index [Select Index Type] • Frequent/One time Scheckuled Indexing Activities [Select Clear Type] • Clear [Select Clear [.*					
arch Engine Setting arch Engines >> Demo Search Engine Basic Settings Advanced Search Features Indexing checkule Setting ndex [Selt Index Type] Frequent/One time Scheduled Indexing Activities Index [Selt Coptimize Type] Chear [Selt Coptimize Type] Schedule now eduled Activities				,		
arch Engine Setting				<u>`</u>		
Basic Settings Advanced Search Features Basic Settings Advanced Search Features checkule Setting Indexing Indexing checkule Setting Frequent/One time Scheduled Indexing Activities imize [search Growthew Control of Schedule now	rch Engine Setting					
Basic Settings Advanced Search Features Indexing checkule Setting index [Sekt index Type] fresh [Sekt Refresh Type] Clear [Sekt CortType] Scheckule now eduled Activities Scheckule now	arch Engines >> Demo Search Engine					
eheckule Setting index [set d index.Type] fresh [set d fereh Type] Frequent/One time Scheduled Indexing Activities Frequent/One time Scheduled Indexing Activities Schedule now eduled Activities Schedule now	Basic Settings	Advanced	Search Features Indexin	g		
ndex [Select index Type] Frequent/One time Scheduled Indexing Activities Frequent/One time Scheduled Indexing Activities Frequent/One time Scheduled Indexing Activities Schedule now Schedule now Schedule now Schedule now	hedule Setting					
Back Link Ck Mark (type) Immediate Link (type) Immediate Link Ck Mark (type) Immediate Link (type) Immediate Link (type) Immediate Link (type) Clear (type) Immediate Link (type) Immediate Link (type) Immediate Link (type)	Idex (Selact Index Type)					
timize [Seka Optimize Type] • Clear [Seka Clear Type] • cx Skyle Schedule now Schedule now	resh (Select Refresh Type)	Frequent/One til	me Scheduled Indexing Activities			
Clear (Sekt Clear Type) C	mize (Select Optimize Type) 🔹					
Schedule now	lear (Select Clear Type)					
Reduled Activities	SAVE	<u> </u>	Schedule now			
E SELECTED SCHEDULE NOW REFRESH	eduled Activities			1		

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.

Schedu	le Setting		
Index	Daily	▼ 06/29/2015 11:00 pm	
Refresh	(Select Refresh Type)	T	
Optimize	(Select Optimize Type)	T	
Clear	Weekly	▼ 06/29/2015 12:30 am	

b. Setup "schedule now" indexing activities.



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

Refresh (Se lect Refresh Type) Optimize (Se lect Optimize Type) Clear Weekly BACK SAVE Schedule Now Activity Type *	
Optimize (Select Optimize Type) Clear Weekly O6/29/2015 12:30 am Schedule Now Activity Type *	
Clear Weekly O6/29/2015 12:30 am Schedule Now Activity Type *	
BACK SAVE Schedule Now	
Activity Type	×
Scheduled Activities Index LINE SCHEDULE NOW REFRESH SAVE CANCEL	

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.
- 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.