Generating Sitemap with SOLR - SOLR Module

The alternative to deprecated Sitemap Module is a Solr-based Sitemap as Indicated on the Deprecations Page:

As Magnolia allows Solr-based searches, we recommend using a Solr-based sitemap instead.

- Step-by-step integrating Solr in Magnolia and generating custom sitemap (a Magnolia Community Wiki page).
- Kevin Cowan: Generating a Sitemap from a Solr Index.

In addition to Step-by-Step integration Solr in Magnolia, you can use the solr_module. It lets you use a template in order to generate a sitemap. the template uses a special templating function: solrfn, which it´s included in the module

In git repository https://git.magnolia-cms.com/projects/SERVICES/repos/solr_sitemap/ there is a module with the necessary templates and class to generate a Sitemap based in SOLR. Module has the following contents:

- Java class: SiteMapTemplatingFunction.java with the templating function necessary to get all Solr Indexes
- Decorations folder: to modify Travel Demo templates and dialogs to allow hide pages in Sitemap.
- Templates folder: with Sitemap template to use directly in Pages App.

Module Installation:
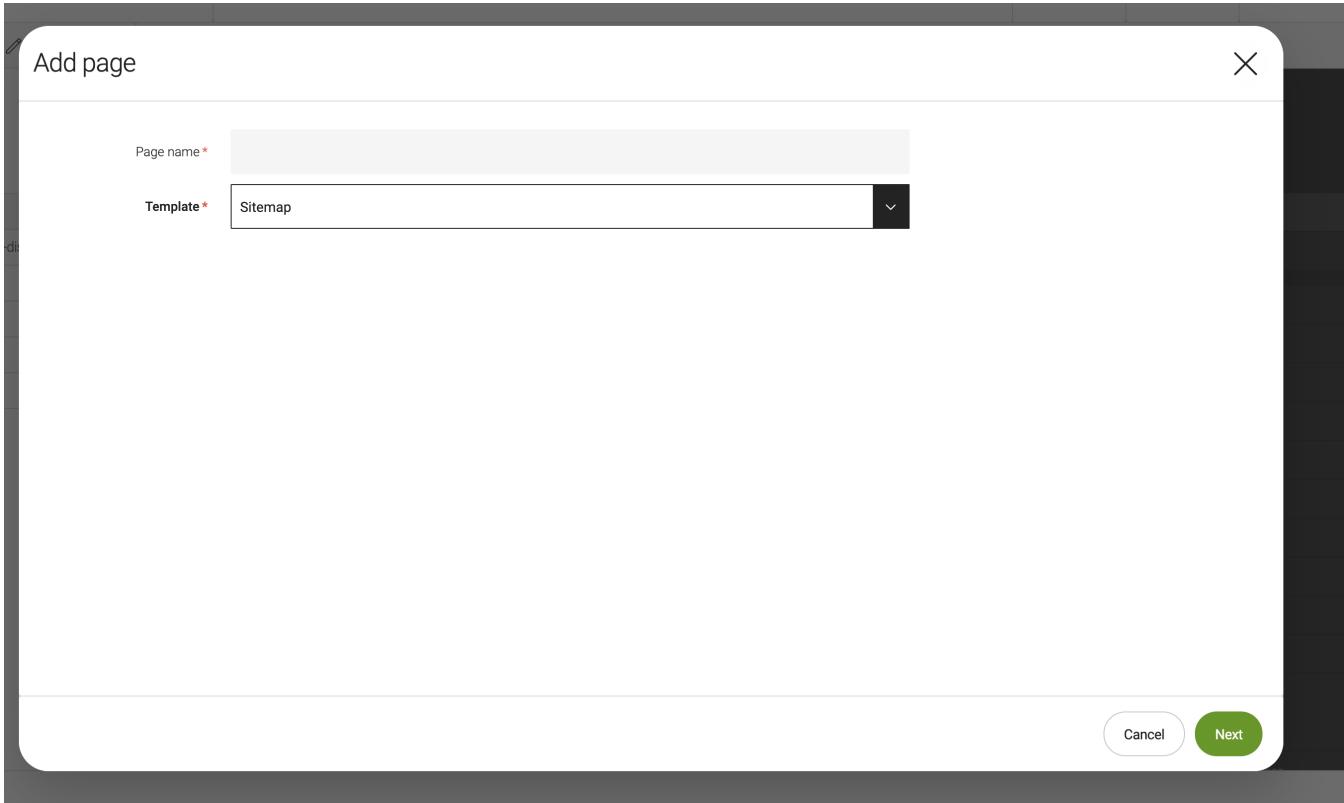
This module is not located in Magnolia Nexus, but you can clone the git repository, generate the jar file and install it in your bundle. Then you need to configure the crawler in content-indexer module after configuring Solr as described in the Step by Step guide

Usage of the module:

You need to register the templating function as shown bellow:

⌄ 🗀 rendering		●
⌄ 🗀 renderers		●
⌄ ◇◇ freemarker		●
⌄ ◇◇ contextAttributes		●
› ◇◇ cms		●
› ◇◇ cmsfn		●
✔ ⌄ ◇◇ solrfn		●
◇ componentClass	test.SiteMapTemplatingFunction	String
◇ name	solrfn	String

Create the Sitemap Page using sitemap template from the module:

Add page

Page name *

Template * | Sitemap

Cancel   Next

Hide Pages in the sitemap:

We create a checkbox in Page dialog to allow endusers decide to show or hide a Page in Sitemap. This check, called "Hide In SiteMap", includes a metatag in the template to avoid indexing by crawler. This tag is <meta name="robots" content="noindex">.

Page properties

Main	Meta Data	Search	Cookie Consent Wording	Cookie Consent Styling

Page title (en) | Travel Home | ?

Navigation title (en) | | ?

Window title (en) | | ?

Navigation | ☐ Don't show this page in any site navigation | ?

Hide In SiteMap | ☐ Hide

Crawler:

You can start the crawler with a cron expression (every hour) or by publication, when you publish a page.

Is possible to delete all indexed elements in Solr, to do this you need a Terminal window:

```
curl http://localhost:8983/solr/magnolia/update --data '<delete><query>*:*</query></delete>' -H 'Content-type:
text/xml; charset=utf-8'
curl http://localhost:8983/solr/magnolia/update --data '<commit/>' -H 'Content-type:text/xml; charset=utf-8'
```