

Google Sitemap XML Generator - <http://googlesitemapxmlgenerator.riaforge.org/>

Google Sitemap XML Generator creates the XML file for the Sitemap protocol (<http://www.sitemaps.org>) used by several search engines such as Google, MSN, and Yahoo. It spiders to create site navigation tree in query format or xml sitemap format.

The methods are:

- **googleSitemap**
- **indexIt**
- **sitemap**
- **scan** (private)
- **submitSitemap**

The code was originally built by Merlinox. (merlinox at gmail dot com .) <http://tinyurl.com/26lzs3>

Download Project - <http://googlesitemapxmlgenerator.riaforge.org/index.cfm?event=action.download>

GoogleSitemapXMLGenerator.zip contains files:

- GoogleSitemapXMLGenerator.doc – for most recent please refer to the Subversion repository
- _sitemap.cfc – main cfc
- sitemap.cfm – example of all methods

Requirements -Should work with Coldfusion MX+

Methods

googleSitemap - remote string googleSitemap (required string root, required query query)

Google Sitemap creation from sitemap result query Page who called "googleSitemap" functon may prepend:

Output: enabled

Parameters: **root**: string, required, root - Root of analyzed site: http + domain + path (es.: <http://www.example.com/site1/page>) **query**: query, required, query

- Result query: it needs "page" data column

Return: string: string with formatted xml (google sitemap)

indexIt - remote query indexIt (required string root, required query query, numeric maxrows="500")

Spider who builds query with pages name and all texts in plain/text format starting from navigation tree query (page column) Result query may insert on a verity collection to create search engine data source

Output: suppressed

Parameters: **root**: string, required, root - Root of analyzed site: http + domain + path (es.: <http://www.example.com/site1/page>) **query**: query, required, query

- Result query: it needs "page" data column maxrows: numeric, optional, maxrows - Max number of indexed pages

Return: query: query (date, page, text)

scan* private void scan (required string root, required string url, required numeric depth)

Scan page and extract all html links (...)

Output: suppressed

Parameters: **root**: string, required, root - Root of analyzed site: http + domain + path (es.: http://www.example.com/site1/page) **url**: string, required, url - Start page of site (es.: index.cfm) **depth**: numeric, required, depth - Click number (depth) of the page from start page

Return: void

sitemap remote query sitemap (required string root, required string url, required numeric depthMax)

*Start spider to scan all pages from start page to all link reachable with *depthMax* click number*

Output: suppressed

Parameters: **root**: string, required, root - Root of analyzed site: http + domain + path (es.: http://www.example.com/site1/page) **url**: string, required, url - Start page of site (es.: index.cfm) **depthMax**: numeric, required, depthMax - Max click number (depth) of the page from start page

Return: query: query (page, depth, analyzed) - analyzed is a boolean

submitSite

Submits site to MSN, Yahoo, Google, and Ask.com

Output: suppressed

Parameters: **url**: string, required, url – location to sitemap.xml (es.: http://www.example.com/sitemap.xml)

Return: return

Method

description

Output: output

Parameters: **parameters**: explanation

Return: return

Example from sitemap.cfm

```
<cfset ini_root="http://www.example.com/machblog" />
<cfset ini_startPage="index.cfm" />
<cfset ini_levelMax=3 />
<cfset ini_sitemap="http://www.example.com/sitemap.xml" />
```

```
<cfinvoke component="example.path.to.cfc._sitemap" method="sitemap" returnvariable="qry_sitemap" root="#ini_root#" url="#ini_startPage#"
depthMax="#ini_levelMax#">
```

```
<cfinvoke component="example.path.to.cfc._sitemap" method="indexIt" returnvariable="qry_sitemapIndex" root="#ini_root#"
query="#qry_sitemap#">
```

```
<cfinvoke component="example.path.to.cfc._sitemap" method="googleSitemap" returnvariable="sitemap" root="#ini_root#">
```

```
query="#qry_sitemap#">
    <cfinvoke component="example.path.to.cfc._sitemap" method="submitSitemap" url="#ini_sitemap#" returnvariable="qry_sitemap" >

    <cffile
    action = "write"
    file = "#ini_sitemap#"
    output = "#sitemap#"
    addNewLine = "no"
    charset = "utf-8">
```

Being the method it declares to you as “remote” he is usable also as web service, important to declare timeout high for situated complexes and scansions to elevated depth (depth > 3).

Issue Tracker

Please enter any issues for this project here: <http://googlesitemapxmlgenerator.riaforge.org/index.cfm?event=page.addissue>

Subversion Access

You may access this project's Subversion repository with your client here: <http://svn.riaforge.org/GoogleSitemapXMLGenerator>

To view this repository with your browser, go here: <http://GoogleSitemapXMLGenerator.riaforge.org/index.cfm?event=page.svnbrowse>

Anonymous users have read access to the repository while the administrator has write access.

This project is sharing its code via Subversion. Subversion is an open source source control method. You may find more information about about Subversion here: <http://subversion.tigris.org/>

Additional Information

RSS Feed -

<http://www.henke.ws/machblog/index.cfm?event=showCategoryRSS&categoryId=DD059F0E-188B-4E84-1516F7AC3CB6BD0B&categoryName=Google%20Sitemap%20XML%20Generator>

Blog -

<http://www.henke.ws/machblog/index.cfm?event=showEntriesByCategory&categoryId=DD059F0E-188B-4E84-1516F7AC3CB6BD0B&categoryName=Google%20Sitemap%20XML%20Generator>

Note: Should add “Sitemap: <sitemap_location>” to your robots.txt file also