

Table of Contents

Introduction	2
Manage Content Quality.....	2
Website Monitoring.....	2
Web Page Monitor:.....	2
Server File Monitoring:	2
Ping Monitor:.....	2
Custom Monitor:	2
Other Features	2
FTP Explorer & Scheduled Download.....	2
sX Web Browser.....	2
Alerts Notifications	3
Offline Analysis	3
Export/Import.....	3
Getting Started.....	3
Scanning Website Contents:	4
Link Explorer.....	5
Identify Broken Links.....	5
Check Email Address, External Links & Resources.....	6
SEO & Optimization Hints	7
Website & Server Monitoring	8
Monitoring Dashboard.....	8
How to Define Monitoring Job	8
Example – Create Monitoring Job to Ping Server	8
Execution Summary – Multiple Jobs:	9
Alert Log (Failure Instances):	10
System Tray Notification:	11
Web Page Monitor	11
Server Files Monitoring.....	11
Generate Meta File	11
Create Server Monitoring Job.....	12
Custom Monitor	13
Scheduled (FTP) Download	13
Generate, Upload and Submit Sitemap	14
Tools - sX Web Browser	15

Last Updated On: Nov, 24 2012

Introduction

SiteXposure provides capabilities to validate quality of website contents and also provides website monitoring features. The combination of Monitoring jobs not only provides statistics of website & server availability but also monitors changes in website contents that may occur due to website hacking or due to unexpected mistakes.

Manage Content Quality

- Website crawler scans all pages on website as per user defined criteria.
- **Verify broken links**
- **Linking to Malicious Sites** : The domain linking allows verifying contents linked with external domains that you may use to verify that your website does not have links or contents (Scripts/CSS/Images) from malicious sites.
- **Identify Security issues**: The webpage metadata and page header is scanned for security issues and recommendations are provided as per HTML best practice guidelines.
- **SEO (Search Engine Optimization) Hints**
- Easily identify **Email Address** & direct **File Links** provided on your website.
- Generate XML **Sitemap** and submit on search engines.

Website Monitoring

Following type of monitoring jobs may be defined to run on user defined schedules.

Web Page Monitor:

- Reports the availability statistics of webpage(s).
- Scan Webpage modification for abusing keywords.
- Multiple options to define your own success/failure criteria. Alert is raised if a success criterion fails or if failure condition is realized.

Server File Monitoring:

- Uses a baseline configuration to monitor changes on Server side files and raise Alerts if any change is identified in comparison to baseline configuration.
- Monitor addition/removal of files or any changes made in existing files on server side.

Ping Monitor:

- Check Server availability by ping requests and raise Alerts if server is unavailable.

Custom Monitor:

- Use Customer Monitor to run an external program. The success criteria is defined based on user inputs that is compared with the final response output captured from external program execution. If success criteria is not met then Alert is raised.

Other Features

FTP Explorer & Scheduled Download

The FTP explorer provides the basic FTP functionality. You may schedule FTP downloads for regular backup of website contents.

sX Web Browser

The web browser provides greater insight of active Web Page including Hidden Fields, Cookies, raw text and details of all HTML tags. Also provides other options like Enable all disabled HTML elements, change styles and invoke Java

Script functions. You shall use these features to evaluate hidden aspects of web pages that can be easily manipulated by other users using similar tools.

Alerts Notifications

The Monitoring dashboard provide list of all Alerts raised for specific Job. In case application is minimized then real time alert is displayed with system tray notification.

Offline Analysis

You may save scan results in a local file and you may export Optimization Hints in CSV formant to analyze with other tools.

Export/Import

Export monitoring job configuration when application is closed and import Monitors configuration to resume monitoring without redefining all jobs.

Getting Started

For website content validation it is necessary to run the website crawler but the default installation comes with a sample configuration file "SampleData.xml" that is installed in the application folder. If you want to just get an overview of features then instead of starting website crawling, you may select the sample configuration file and continue. The "First Time user" section provides the option to open sample configuration file.

Getting Started

First Time User?

- ▶ [Click here](#) to get a brief of how indexing process works.
- ▶ The standard installation contains a sample data file. [Click here](#) to launch Dashboard with sample data file.

▶ To use an existing site configurating file, [click here](#) 📁

▶ To start with a new website, enter website URL in below section and click on Start Indexing

Enter URL of your website to start Indexing :

Indexing Configuration

Max. Link Depth:

Max. instances of URL:

Max. links to crawl:

Request Timeout (seconds):

☒ Verify Links to External Domain

☐ Allow Blocked Domain

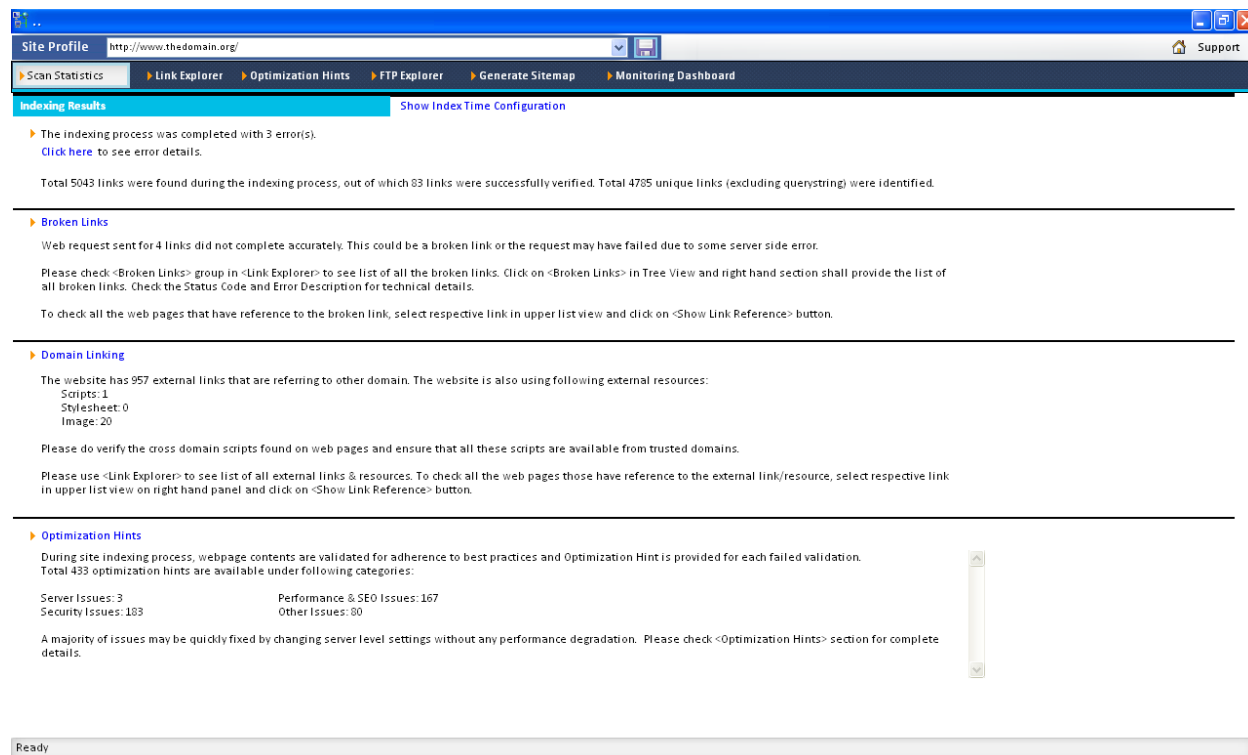
[Configuration Help](#)

▶ [Click here](#) to change default Content Exclusion Criteria

Start Indexing

Ready

Scan Statistics - Dashboard



Once the web site crawling is finished or when you open existing/sample data file, the scan statistics dashboard is display that provides the summary of all the issues identified during site scan. The “**Link Explorer & Optimization Hints**” tabs provide detailed insight of website contents and issues identified with respective webpages. The “**How It Works**” option provided on each tab provides more details about how to use individual features.

Scanning Website Contents:

Following section provide the overview of configuration parameters that define and restrict the indexing process. Once the scan is completed you may save indexing results for offline analysis on later stage.

Indexing Parameters:

Verify Links to External Domains: If this flag is on then during Site indexing process, all external links will be verified so that broken links may be reported.

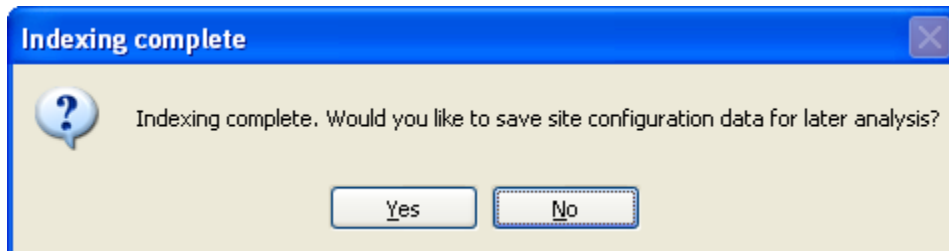
Max. Links to Crawl: Defines the limit of maximum page requests that would be sent to the webserver and when maximum request limit is reached the indexing process shall terminate. However all the processed results would be retained for analysis.

Max. Link Depth: Defines the nested level for the page links to be crawled. For example, to scan only first/landing page enter 0, to further scan nested link available on landing page enter 1 and so on.

Max. Instances of URL: The website may have multiple instances of same page differentiated with just query string parameters. There could be dynamic pages as in the case of Calendar based interfaces and this may result in infinite loop. The system will only hit a URL for the maximum instances as defined with this parameters and rest of the links would be ignored.

Stop Indexing: You may terminate the Indexing process at any stage by clicking on the **Stop Indexing** button. The indexing process will terminate. However all the processed results would be retained for analysis.

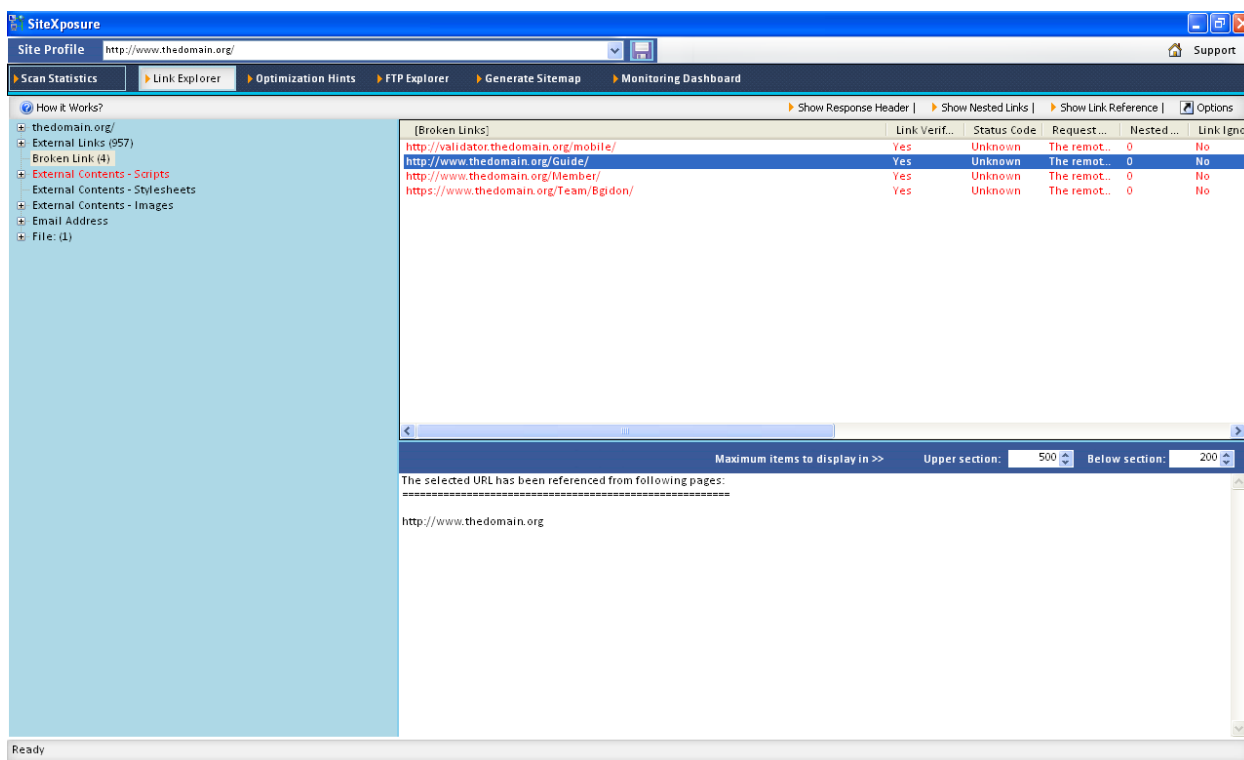
When the indexing process is completed, system will prompt to save the indexing results into a configuration file on local disk. Please do save the configuration details in case you would like to retain all analysis data for offline access.



Link Explorer

Identify Broken Links

The **Link Explorer** tab provides all the details about your site links.



Select **Link Explorer** tab and click on **Broken Links** item in left tree view. The right hand section will be loaded with all the broken links found on website. To check all the **references of a Broken Link**, select the link in right hand upper section and click on **Show Link Reference**. This will show the list of all webpages that have reference to respective broken link.

Other Options on Link Explorer Tab:

To check all the links defined on a webpage, select the link in right hand upper section and click on **Show Nested Links**.

Include Skipped/Ignore Links Option: The indexing process has an inbuilt exclusion criteria for example it does not scan pictures/ images/ tag pages etc. and all such links are ignored during indexing process. To hide such links from link list, you may disable this option.

The **Show Page Header** button will display the page header information as below:

X-Pingback: <http://www.domain.com/xmlrpc.php>

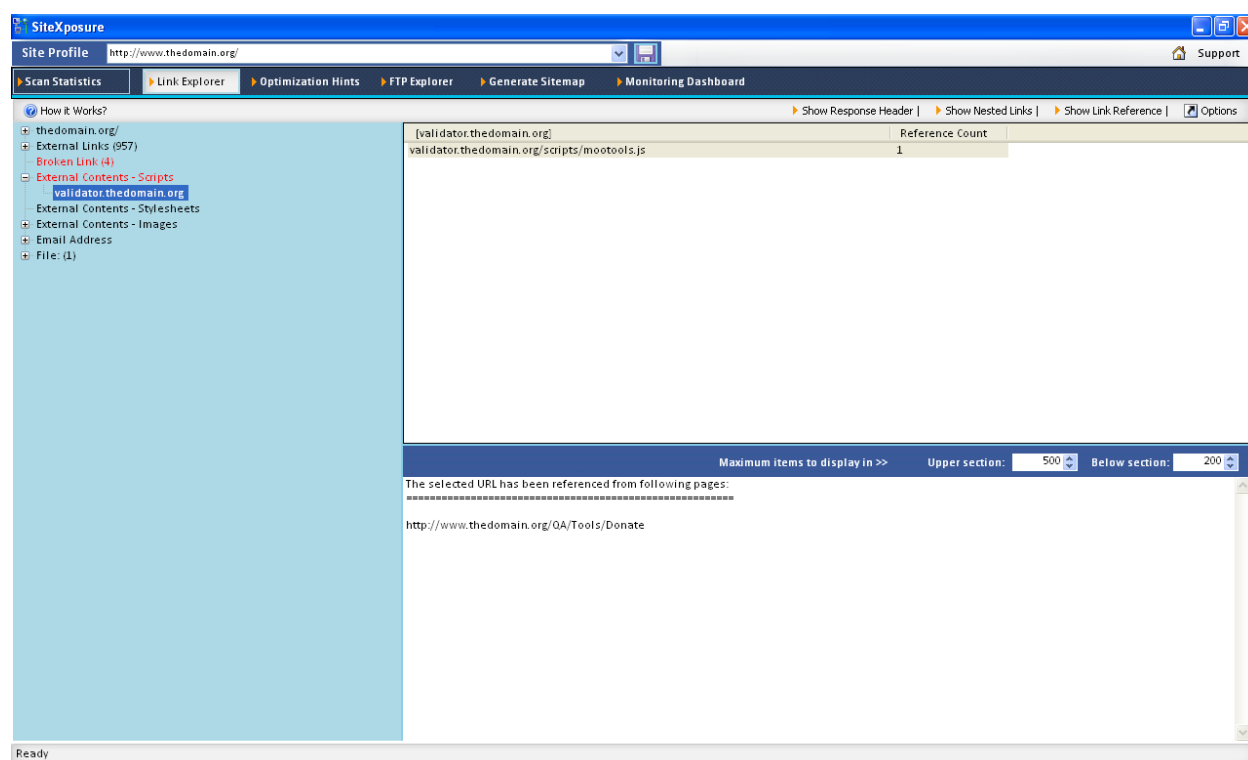
Vary: Accept-Encoding

Transfer-Encoding: chunked

Content-Type: text/html; charset=UTF-8

Date: Mon, 01 Aug 2012 16:41:57 GMT

Check Email Address, External Links & Resources



The Link Explorer also allow you to check all the external links and other resources like media, CSS and Scripts that are linked with external domains.

The Link Explorer also shows the **email address and File links** available on the website.

To check the reference source of each type of resource/external links, use **Show Page Source** option.

SEO & Optimization Hints

SiteXposure scans the webpage HTML & headers to identify gaps as per best practices. The optimization hints tab lists all such issues identified during the scan process.

SiteXposure

Site Profile <http://www.thedomain.org/> Support

Scan Statistics | Link Explorer | **Optimization Hints** | FTP Explorer | Generate Sitemap | Monitoring Dashboard

How it works? Filter by Category: Refresh URL List Export Details More Details (Technical Reference) Options

Category	Rule	Page Count
Server Issues	OPTIONS Method Exposed	NA
Server Issues	Information Disclosure By Extra Header SERVER	NA
Server Issues	Proxy Interception	NA
Other Issues	Charset Not UTF 8	5
Other Issues	Cache Control Not Defined	8
Other Issues	No Email Meta Tag Missing	1
Performance & SEO	Keyword Meta Tag Missing	48
Performance & SEO	Resource Type Meta Tag Missing	1
Security Issues	Windows Target Meta Tag Missing	58
Performance & SEO	Title Tag Too Long	2
Other Issues	Description Tag Missing	53
Security Issues	X Frame Options Header Not Defined	58
Security Issues	MIME Sniffing Not Enabled	58
Performance & SEO	Incorrect URL Structure	3
Performance & SEO	Alt Attribute Missing	83
Performance & SEO	Too High Link On Page	29
Other Issues	DOCTYPE Tag Not Defined	13
Performance & SEO	HTTP Compression Not Enabled	1
Security Issues	Cross Domain Data Transfer	9

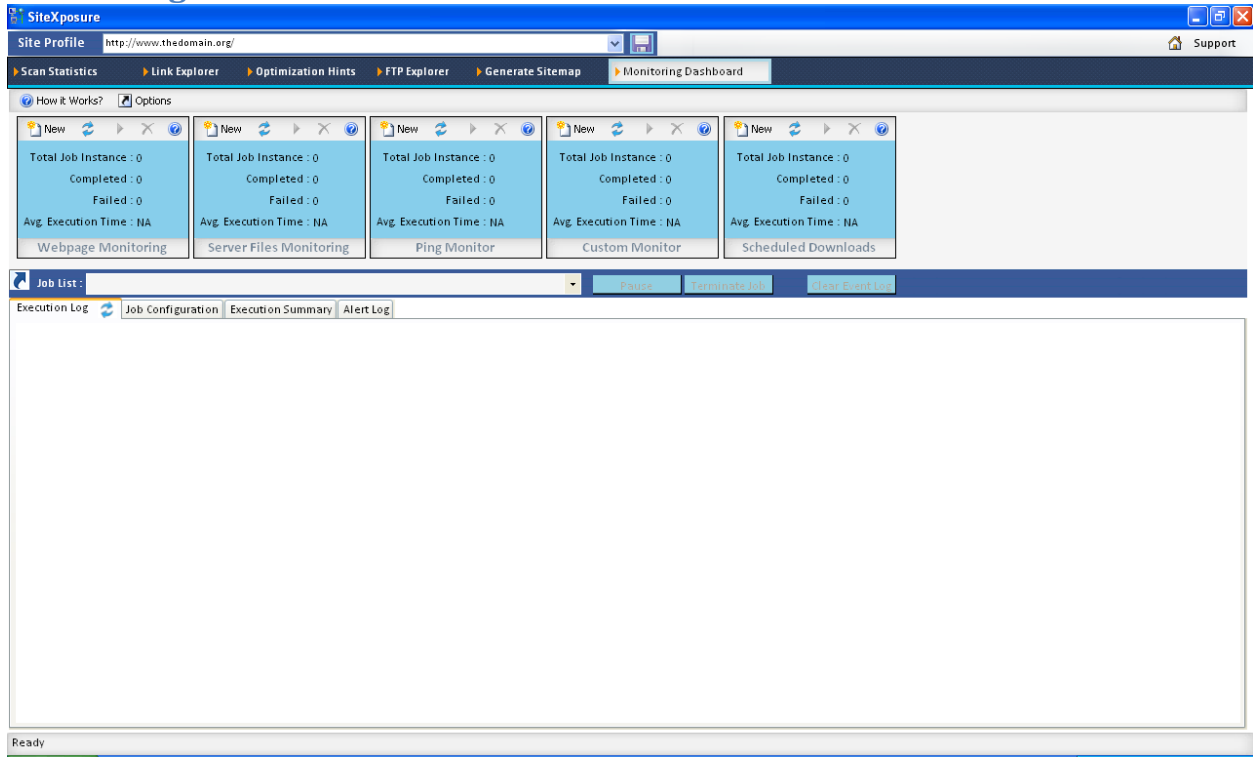
Maximum number of webpages to list in below section: 100

[URL: <http://www.thedomain.org/>]
Extra information disclosed by HTTP header [Server : Apache/2]

Ready

Website & Server Monitoring

Monitoring Dashboard



Features summary:

- Create new monitoring jobs to run in background.
- See real time execution log and other reports for a Job or Job Category.
- Disable/Resume all monitors or selected job or terminate a Job permanently.
- Refresh log history for a job.
- Export monitoring job configuration when application is closed and import Monitors configuration to resume monitoring without redefining all jobs.

How to Define Monitoring Job

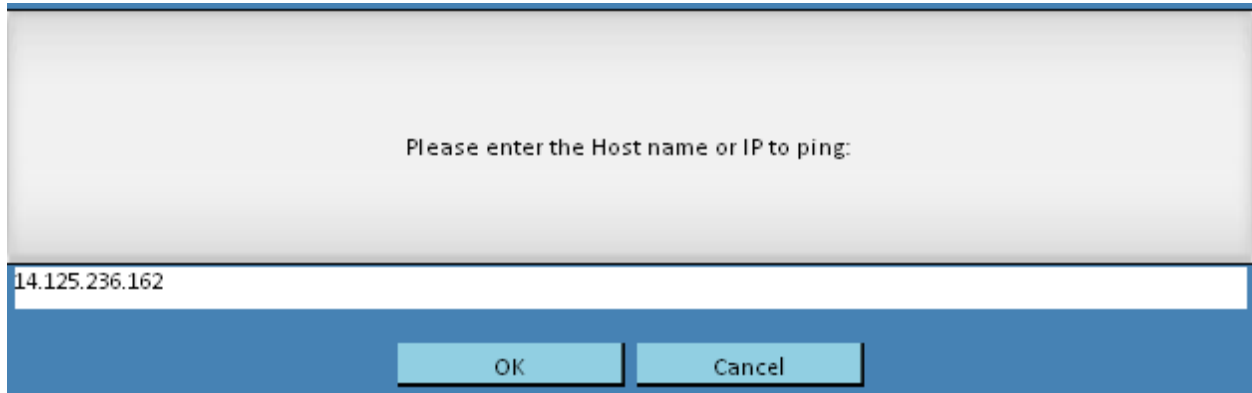
The Ping monitor works with minimum configuration hence explaining the process to create monitoring job & reporting using the Ping Monitor. Same process is applicable to other Monitors.

Example – Create Monitoring Job to Ping Server

Go to Monitoring Dashboard and click on “New” option on “Ping Monitor” box.

Provide server IP/name and define the ping schedule/frequency.

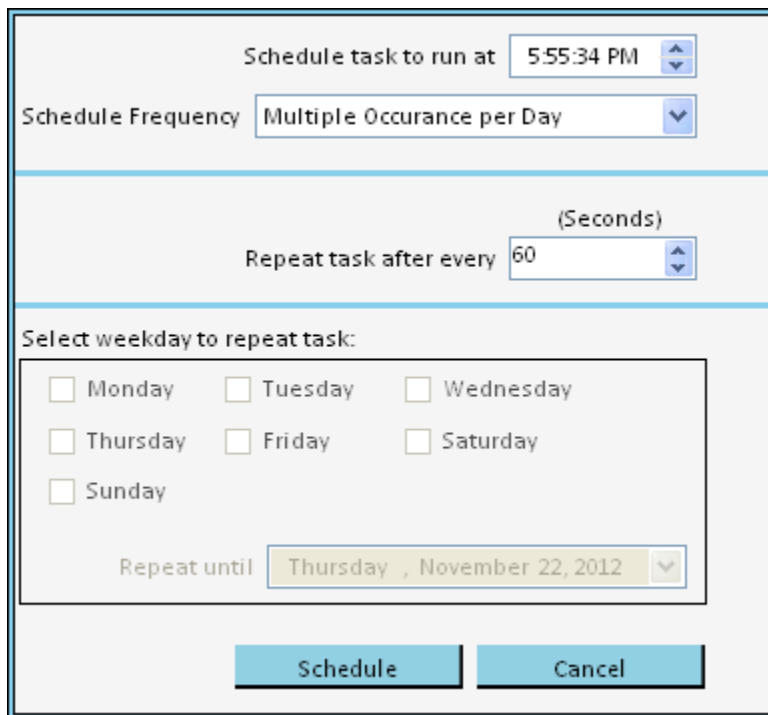
The ping command would be sent to server on defined frequency/schedule and result of ping would be available with Execution log, Execution Summary report and Alert Log.

Enter Server IP or Name:

Please enter the Host name or IP to ping:

14.125.236.162

OK Cancel

Define Ping schedule or Frequency:

Schedule task to run at 5:55:34 PM

Schedule Frequency Multiple Occurance per Day

(Seconds)

Repeat task after every 60

Select weekday to repeat task:

☐ Monday ☐ Tuesday ☐ Wednesday

☐ Thursday ☐ Friday ☐ Saturday

☐ Sunday

Repeat until Thursday , November 22, 2012

Schedule Cancel

Execution Summary – Multiple Jobs:

To see the Execution Summary for all Ping Jobs, please click on “Execution Summary” tab, select “Ping Monitor” and click on Refresh button.

The screenshot shows the SiteXposure Monitoring Dashboard. The top navigation bar includes links for Site Profile, Scan Statistics, Link Explorer, Optimization Hints, FTP Explorer, Generate Sitemap, and Monitoring Dashboard. Below the navigation bar, there are five panels for different monitoring types: Webpage Monitoring, Server Files Monitoring, Ping Monitor, Custom Monitor, and Scheduled Downloads. The Ping Monitor panel is selected, showing a summary of job instances. Below the summary, there is a table with the following data:

Job Instance	Instance Started	Instance Completed	(%)	Instance Failed	(%)	Status unknown	Avg. Execution Time
74.125.236.162 (1)	3	3	100 %	0	0 %	0	75 (ms.)
123.123.41.123 (2)	3	3	100 %	0	0 %	0	163 (ms.)
Test (3)	7	0	0 %	7	100 %	0	0 (ms.)

The bottom status bar shows 'Ready'.

Alert Log (Failure Instances):

To see details of failed Ping against a Server, please select respective Job in “Job List” pane and click on “Alert Log” tab. The Alert Log list shows details of failed instances for the job selected in “Job List” pane.

The screenshot shows the SiteXposure Monitoring Dashboard with the Alert Log tab selected. The top navigation bar and summary panels are the same as in the previous screenshot. The Alert Log tab displays a table with the following data:

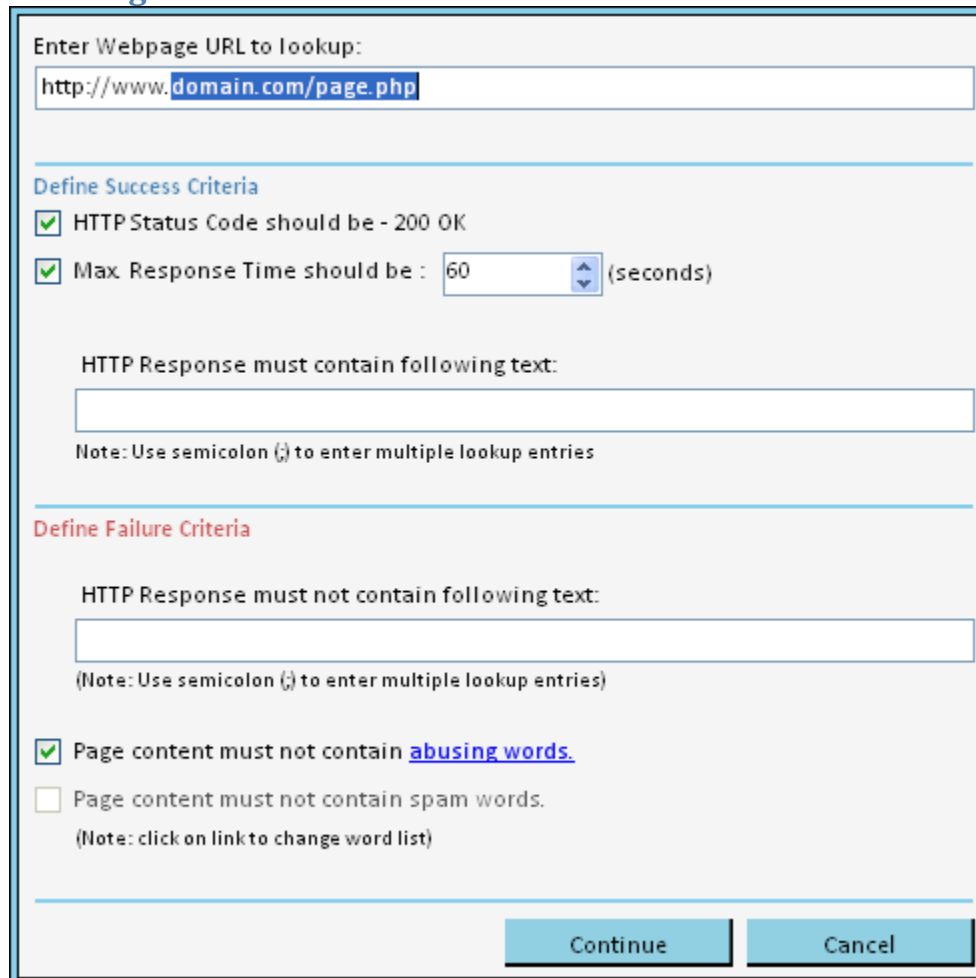
Job Instance	Alert Description	Log Time	Job Execution Time
Test	Ping failed to Test	11/21/2012 6:03:13 PM	0 (ms.)
Test	Ping failed to Test	11/21/2012 6:03:13 PM	0 (ms.)
Test	Ping failed to Test	11/21/2012 6:03:19 PM	0 (ms.)

The bottom status bar shows 'Ready'.

System Tray Notification:

In case SiteXposure is running in minimized mode, then Alert notification is provided on the System Tray.

Web Page Monitor



The screenshot shows a 'Web Page Monitor' configuration window. At the top, there is a text box labeled 'Enter Webpage URL to lookup:' containing the text 'http://www.domain.com/page.php'. Below this is a section titled 'Define Success Criteria' in blue. It contains two checked checkboxes: 'HTTP Status Code should be - 200 OK' and 'Max. Response Time should be : 60 (seconds)', where '60' is in a spin box. Below these is a text box for 'HTTP Response must contain following text:' which is currently empty. A note below it says 'Note: Use semicolon (;) to enter multiple lookup entries'. The next section is 'Define Failure Criteria' in red. It has a text box for 'HTTP Response must not contain following text:' which is also empty, with a note '(Note: Use semicolon (;) to enter multiple lookup entries)'. Below that are two checkboxes: 'Page content must not contain abusing words.' (checked) and 'Page content must not contain spam words.' (unchecked). A note at the bottom of this section says '(Note: click on link to change word list)'. At the bottom right of the dialog are 'Continue' and 'Cancel' buttons.

Features summary:

- Reports the availability statistics of webpage(s).
- Scan Webpage modification for abusing keywords.
- Multiple options to define your own success/failure criteria. Alert is raised if a success criterion fails or if failure condition is realized.

Server Files Monitoring

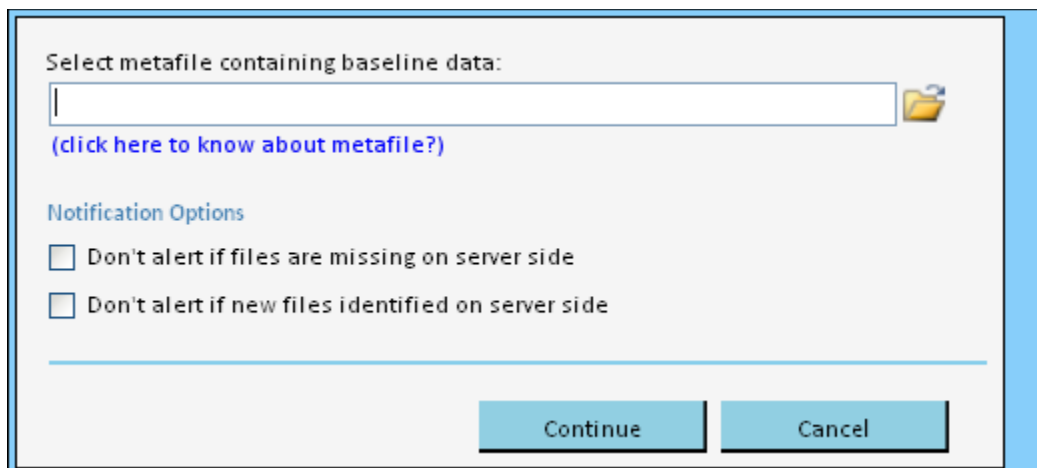
You may create monitoring job to create a baseline server configuration/meta file using FTP protocol. The Meta file can be generated using "FTP Explorer".

Generate Meta File

Once you are connected with FTP server, select “Generate Metadata File” in “Options” menu. This would scan list of all file along with file attributes in selected/working FTP folder and when scanning process is complete, system will prompt to save Meta file on the local file system.

Create Server Monitoring Job

When Meta File has been generated, go to Monitoring Dashboard and click on “New” option on “Server File Monitoring” box. System will prompt to select Meta file from local file system.



Select metafile containing baseline data:



[\(click here to know about metafile?\)](#)

Notification Options

☐ Don't alert if files are missing on server side

☐ Don't alert if new files identified on server side

[Continue](#) [Cancel](#)

Select the Meta file generated using “FTP Explorer” and other Notification/Alert options.

Enter FTP connection details, define monitoring frequency and start the Job.

System would start monitoring FTP server in the background for following events and shall raise Alerts if any change in identified:

1. New files identified on server
2. Any file removed from server
3. Any changes identified in existing file attributes (file size, permissions/owners and last modification date).

Note: For performance optimization & security reasons, actual file contents are not downloaded from the FTP server but only directory listing is retrieved. For security reasons it is recommended to use an FTP account that have limited access on FTP server. The FTP password is not stored on local disk in case you export Monitoring Job configuration.

Custom Monitor

Enter shell command or path of program file to execute:

☐ Supress output window/ run program silently

Define Success Criteria

Program output must contain following text:

Note: Use semicolon (;) to enter multiple lookup entries

Define Failure Criteria

Program output must not contain following text:

Error; Failed; Exception;is not recognized as an internal or external command

(Note: Use semicolon (;) to enter multiple lookup entries)

Continue Cancel

SiteXposure allows you to create recurring jobs with your own program or Windows shell commands.

The program or shell command can be executed as per defined schedule & frequency.

SiteXposure captures the output generated by the program/shell command and raise Alert as per the Success/Failure Criteria defined.

For example, to use Windows PING utility, you may enter PING command along with desired IP address and define the list of words that are expected or not expected from Ping results. If desired result is not received then SiteXposure raises the Alert or shows notification in System Tray (if running in minimize mode).

Scheduled (FTP) Download

You may schedule an FTP Download Job that may be used for regular backups of server files. The Job may be scheduled from the FTP Explorer or from the Monitoring Dashboard but you must be already logged on the FTP server (using FTP Explorer).

Note: For security reasons it is recommended to use an FTP account that have limited access on FTP server.

Generate, Upload and Submit Sitemap

The screenshot shows the SiteXposure web application interface. At the top, there's a navigation bar with links: Site Profile, Scan Statistics, Link Explorer, Optimization Hints, FTP Explorer, Generate Sitemap, and Monitoring Dashboard. The 'Generate Sitemap' link is highlighted. Below the navigation bar, the main content area is titled 'Generate Site Map & Submit to Search Engines'. It contains instructions on how to generate and upload a sitemap. The 'Site Map Options' section has checkboxes for 'Include Last Modification' and 'Exclude Broken Links', both of which are checked. There's a 'Generate Site Map' button. Below this, there's a section for uploading the sitemap via FTP, with fields for 'FTP Server URI', 'FTP Account Details' (Username and Password), and '(Upload as) File name'. The 'Upload File' button is also present. At the bottom, there's a section for submitting the sitemap to search engines, with checkboxes for 'Ask.com', 'Google', and 'Bing'. The 'Submit Sitemap' button is also present.

SiteXposure

Site Profile <http://www.thedomain.org/>

Support

Scan Statistics Link Explorer Optimization Hints FTP Explorer **Generate Sitemap** Monitoring Dashboard

Generate Site Map & Submit to Search Engines

▶ To generate XML sitemap and save in local file click on <Generate Sitemap> button. Once Sitemap is generated, you may upload new sitemap and submit on Search Engines using other options in below sections.

Site Map Options:

☒ Include Last Modification

☒ Exclude Broken Links

Last Sitemap Saved At (Local File Path):

Generate Site Map **Generate Sitemap**

▶ To upload sitemap using FTP: 1) Select the local file to upload 2) enter FTP location, Username and Password and 3) click on <Upload File> button. If there is already an existing file with the same name then old file would be renamed as <filename>.bak

FTP Account Details:

User Name :

Password :

Local file path to upload on FTP:

FTP Server URI (Upload as) File name

Upload File

▶ Once you have uploaded new sitemap on your website, then use this option to send notification to Search Engines about updated sitemap. The search engines may not immediately re-index the website but the updated sitemap notification does improve the probability of sooner indexing by search bots.

Enter complete URL of sitemap, select the Search Engine services to be notified and click on <Submit Sitemap> button

Choose Search Engine:

☐ Ask.com ☒ Google ☒ Bing

Sitemap URI [Format: <http://www.yoursite.com/sitemap.xml>] :

Submit Sitemap **Submit Sitemap**

Ready

If you run complete site scan that would index all site links then you may use this option to export site links in standard XML sitemap format to upload on search engines.

To generate the sitemap in XML format, click on the “Generate Sitemap” button and it will generate an XML file as per Sitemap specification.

To upload sitemap on your website:

Browse the sitemap from local file system as saved in previous step. Enter your FTP server URL, username & password and click on Upload File button. This will upload the sitemap on your website. Any pre-existing file would be renamed for backup purpose.

Submit Sitemap to Search Engines:

Once you have uploaded Sitemap on your website, you may enter the sitemap URL into this section and click on Submit sitemap button. This would submit the sitemap URL to Search engines using Google/Bing web service for sitemap submissions. The system will notify of Success/Failure as per submission acceptance provided by search engine submission request.

Tools - sX Web Browser

The web browser provides greater insight of active Web Page including Hidden Fields, Cookies, raw text and details of all HTML tags. Also provides other options like enable all disabled HTML elements, change styles and invoke Java Script functions. You shall use these features to evaluate hidden aspects of web pages that can be easily manipulated by other users using similar tools.

