

MC+A GSS
Administrators User Manual

July 25, 2009

V 2.5.2



3717 North Ravenswood Avenue
Suite 210
Chicago, IL 60613
www.mcplusa.com

Revision History

Version Number	Author	Date	Description
1.0	Michael Cizmar	09/06/2005	Initial Draft
1.1	Michael Cizmar	09/27/2005	Updated for WSS/SPS web parts
2.1	Michael Cizmar	06/08/2006	Update documentation for new version of software
2.2	Michael Cizmar	07/01/2006	Minor updates removing web part information
2.3	Tom Howard	08/21/2006	Revised from initial customers review
2.3.1	Tom Howard	08/31/2006	Revised Licensing
2.4	Michael Cizmar	03/29/2007	Revised instructions
2.5	Michael Cizmar	04/09/2007	Added Setup
2.5.1	Angelo Hakey	06/24/2009	Revised instructions
2.5.2	Angelo Hakey	07/25/2009	Revised instructions

Table of Contents

1 Executive Summary 5

1.1 Introduction 5

1.2 Support 5

1.3 Sales 5

1.4 System Requirements..... 6

1.4.1 Hardware/Operating Systems 6

1.4.2 Software Requirements 6

1.5 Assumptions 6

2 Components 6

2.1 Installation – Setup 6

2.2 Installation – Manual 7

3 License Key 9

3.1 Entering Your License Key 9

3.2 Navigate to GSS 10

4 Common Configuration issues 11

4.1 ASP.NET 2.0 has not been configured. 11

4.2 make sure links are showing up on the gss.aspx page..... 13

4.3 Check the appliance’s configuration..... 18

Appendix A: About MC+A 21

Dear Administrator,

Thank you for acquiring the MC+A GSS. As part of the Google Enterprise Professional program, MC+A has developed this technology to aid in the discovery and indexing of content for your Google Mini.

We have changed the connector from our initial version, which was a web part. This is now manifested currently as an ASP.NET 2.0 web site application that builds a list of links that is scanned by the web crawler. This change should make it easier to configure. You can use standard .NET security or you can embed the credentials within the web.config file.

This version has been tested in a variety of SharePoint environments and current has potential limitless scalability. It operates independently from your SharePoint environment so it has little impact on your production system.

We have worked hard to provide you with a product that is useful, easily deployable and stable. We continue to incorporate feedback we receive and hope that it exceeds your expectations.

We look forward to your feedback.

Regards,

/Emailed/
Michael Cizmar
michael.cizmar@mcplusa.com
President
MC+A

1 Executive Summary

1.1 Introduction

MC+A developed the GSS to allow the Google Mini to index content that resides in Microsoft SharePoint™ Portal and Microsoft SharePoint™ Services. The Google Crawler has difficulty with some aspects of the DHTML of the SharePoint system. This web part displays all relevant content in the form of links so that the crawler can index the system. There is a web part for SharePoint Portal based sites and a web part for Windows SharePoint Services based site. This manual serves as a single source for administrators wishing to install and configure the Google SharePoint Site Map.

The GSS consists of a single ASP.Net 2.0 website. It uses standard Microsoft security and .Net framework technologies.

1.2 Support

Support cases are available for purchase or via the annual support and maintenance plan. Should you require additional customizations or premium support please contact:

MC+A Support
3717 North Ravenswood Avenue
Suite 210
Chicago, IL 60613
312.275.5737 x2
support@mcplusa.com

1.3 Sales

MC+A provides professional services utilizing the Google Search Appliance, Google Mini and Microsoft SharePoint related technologies. Please contact sales if you have specific requirements that are not being met and seek additional services.

MC+A Sales
3717 North Ravenswood Avenue
Suite 210
Chicago, IL 60613
312.275.5737 x1
sales@mcplusa.com
www.mcplusa.com

1.4 System Requirements

1.4.1 Hardware/Operating Systems

In addition to the Google Mini itself, you will need a system that runs on Microsoft Windows Server 2003 with Microsoft .Net Framework 2.0. All systems must be configured to Microsoft specifications.

1.4.2 Software Requirements

No additional software is required.

1.5 Assumptions

This manual will guide the SharePoint system administrator in configuring the system. It is assumed that this person is familiar with the technologies and terminologies associated with SharePoint on the Microsoft platform which includes:

- SharePoint Portal Sever (or Windows SharePoint Services)
- .Net Assemblies
- IIS Web Server Configuration
- Prior to installation of MC+A Google Site Map Web Part, Administrators will follow Google's instructions on configuration of the SharePoint server for the Google crawler (creation of Google user, modifying of templates, etc.)
http://code.google.com/apis/searchappliance/documentation/connectors/110/connector_admin/sharepoint_connector.html

2 Components

2.1 Installation – Setup

1. Run the setup.msi program (contained in the setup folder of the zip). This will:
 - a. Unpack the software to the specified location
 - b. Create a Local User
 - c. Create an IIS Web Application on port 7070
2. Change the following web.config settings to match your environment:
 - a. SiteUrl - with the base URL of the SharePoint site (no trailing "/", i.e. http://mySharePoint.mycompany.com)
 - b. ListPage – for large libraries and lists, the site map will paginate (Default is 100)
 - c. Version - Ignore

3. Configure authentication model to corporate standards
 - a. By default the application sends remote requests as the logged in user by using impersonation.
 - b. UseLocalCred – Literal true or false. We had customers who did not have Kerberos set up correctly. When this is set to true, the GSS attempts to make a Network Credential from the username, password and domain stored in the web.config.
 - c. UserName – The NT User name to make the SharePoint request
 - d. Password – The password, stored as clear text.
 - e. Domain – The Active Directory Domain
4. Navigate to base URL. (<http://localhost:7070/>) Be sure the firewall port 7070 is open to incoming requests.
5. Click GSS link from default page.
6. Confirm a list of links
7. Add the URL of the GSS (i.e. <http://mywebserver.company.local:7070/>) to the “Start Crawling from the Following URLs:”
8. Add the URL pattern of the GSS (i.e. <http://mywebserver.company.local:7070/>) to the “Follow and Crawl Only URLs with the following Patterns”.
9. Add the URL pattern of the SharePoint site (i.e. <http://mysharepointsite.company.local/>) to the “Follow and Crawl Only URLs with the following Patterns”.
10. In 30-40 minutes you should see results in your indexed (assuming your appliance is configured for continuous crawling)

2.2 Installation – Manual

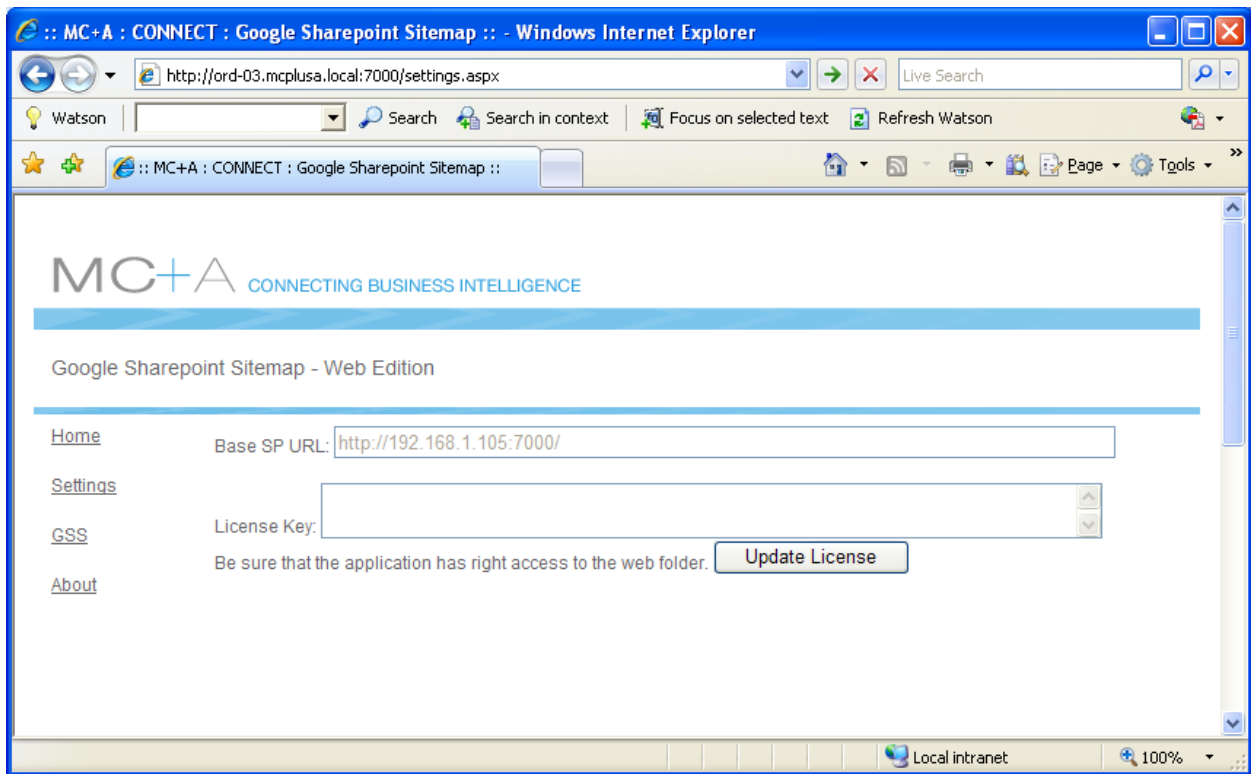
1. Unpack the zip and install the virtual directory.
 - a. Unpack the GoogleSharepointSitemap.site
 - b. Register a new website in IIS
(<http://www.microsoft.com/technet/prodtechnol/WindowsServer2003/Library/IIS/9b74f316071b-4eab-b5d2-c3ad7feb46ca.mspx?mfr=true>)
 - c. Update the path of the new IIS website to where the GoogleSharepointSitemap.site folder is
 - d. Change the default user to a NT user that one has access to read the registry (required for the licensing manager)
 - e. Test the website by launching Internet Explorer and verifying the default web page loads.

2. Change the following web.config settings to match your environment:
 - a. SiteUrl - with the base URL of the SharePoint site (no trailing “/”, i.e. <http://mySharePoint.mycompany.com>)
 - b. ListPage – for large libraries and lists, the site map will paginate (Default is 100)
 - c. Version - Ignore
3. Configure authentication model to corporate standards
 - a. By default the application sends remote requests as the logged in user by using impersonation.
 - b. UseLocalCred – Literal true or false. We had customers who did not have Kerberos set up correctly. When this is set to true, the GSS attempts to make a Network Credential from the username, password and domain stored in the web.config.
 - c. UserName – The NT User name to make the SharePoint request
 - d. Password – The password, stored as clear text.
 - e. Domain – The Active Directory Domain
4. Navigate to base URL. (<http://myapp.mycompany.com/gss>)
5. Click GSS link from default page.
6. Confirm a list of links
7. Add the URL of the GSS (i.e. <http://mywebserver.company.local:7070/>) to the “Start Crawling from the Following URLs:”
8. Add the URL pattern of the GSS (i.e. <http://mywebserver.company.local:7070/>) to the “Follow and Crawl Only URLs with the following Patterns”.
9. Add the URL pattern of the SharePoint site (i.e. <http://mysharepointsite.company.local/>) to the “Follow and Crawl Only URLs with the following Patterns”.
10. Go to Crawler Access and add the URL of SharePoint site, the domain that it is in, and the same set of credentials that is used for the GSS.
11. In 30-40 minutes you should see results in your indexed (assuming your appliance is configured for continuous crawling)

3 License Key

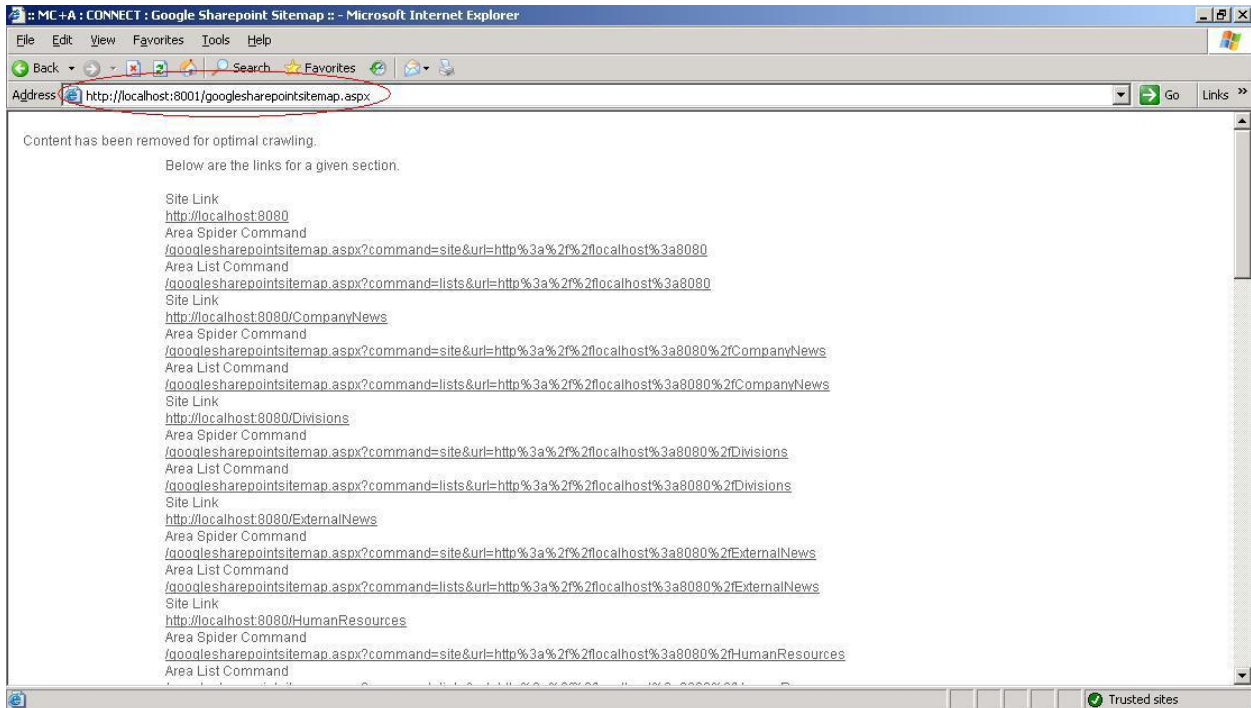
3.1 Entering Your License Key

The software by default runs for 30 days without a key. Past that time, you must obtain a valid license from MC+A. Your license is unique to the server and your order. You will simply save the license file to the root of the web folder when you obtain it.



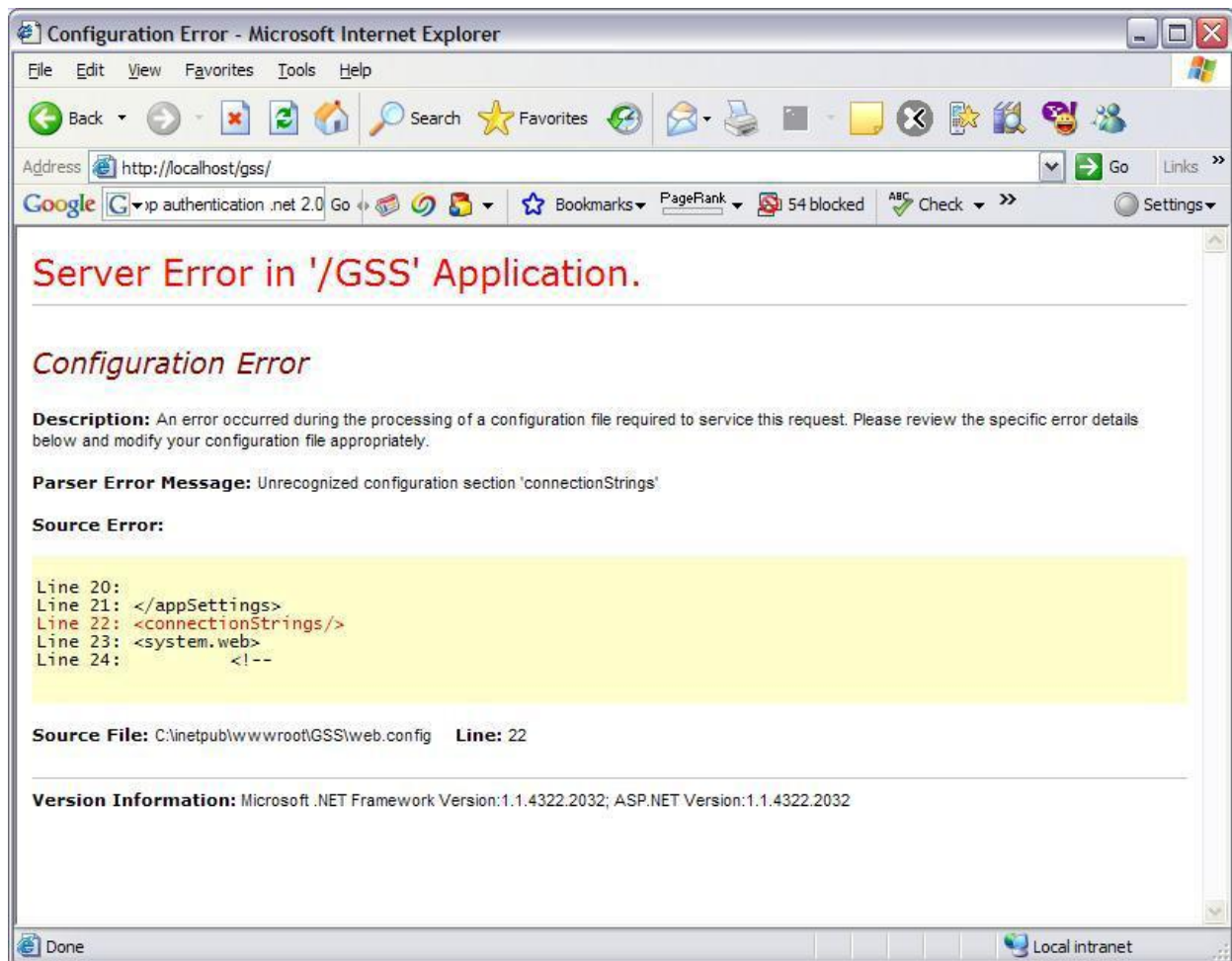
3.2 Navigate to GSS

Go to the URL of the GSS and click on the GSS link. If you do not see an exception or other error you have successfully installed the GSS. If you have any problems, email: support@mcplusa.com



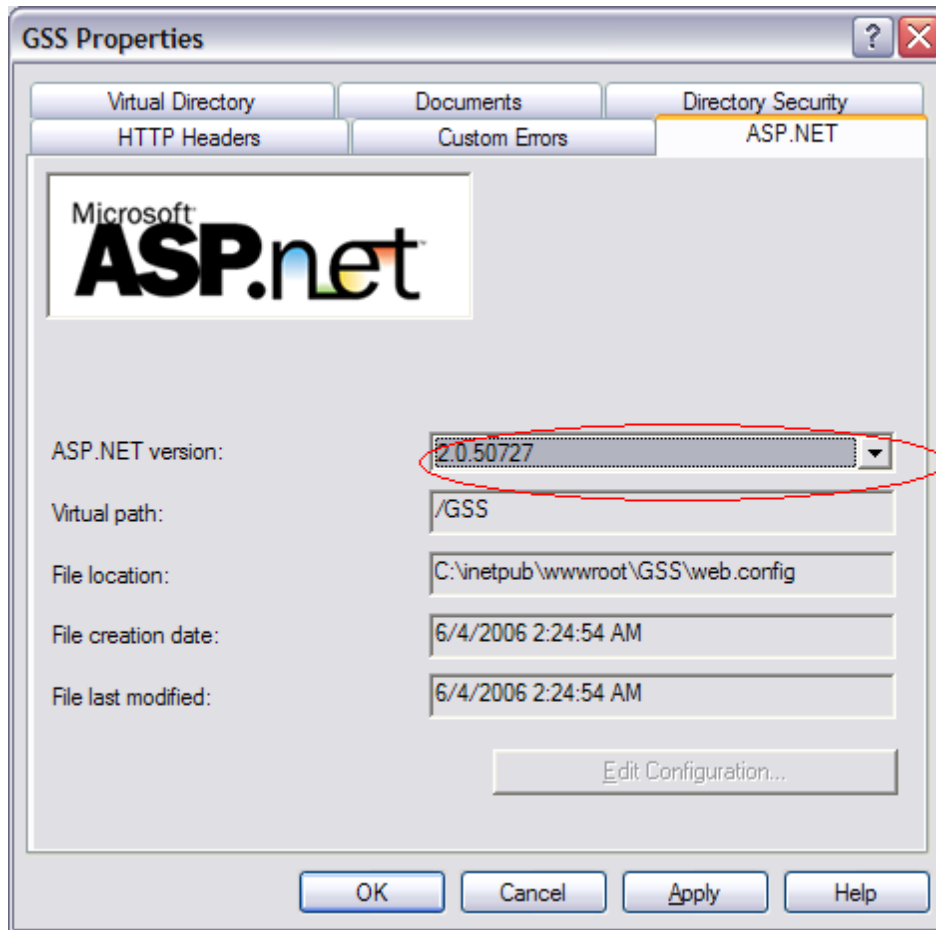
4 Common Configuration issues

4.1 ASP.NET 2.0 has not been configured



This project requires ASP.NET 2.0 If your default virtual directory is ASP.NET 1.1 this message will appear when you navigate to the site.

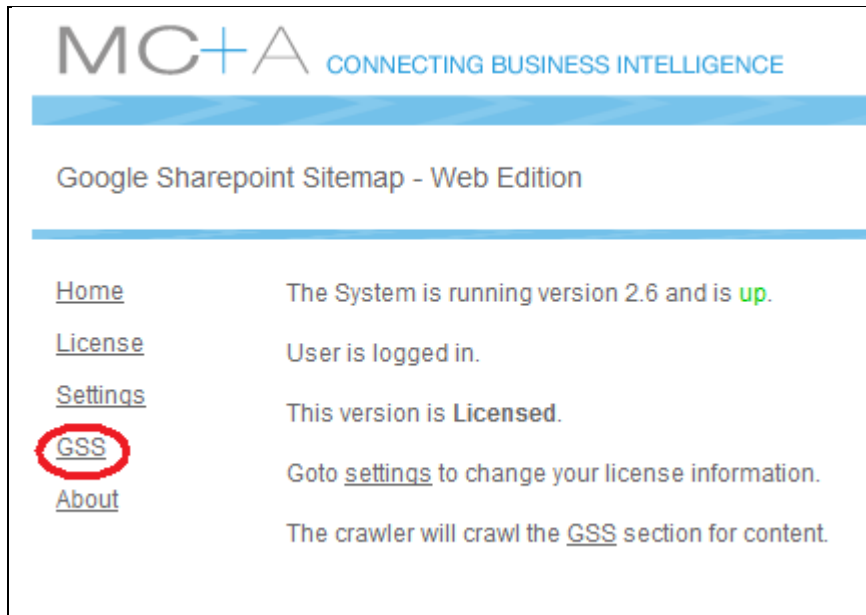
Simply go to IIS Manager and change the following setting for your web site:



4.2 make sure links are showing up on the gss.aspx page

Often time's customers immediately assume that something is wrong with their mini when in fact they don't have their Google Sitemap properly configured. Check to see if the links in the sitemap are working.

Click on the **GSS** link



Now click on various links to ensure that there are no HTML errors. Exclude the **Site Link** entries from this test as their results will not be relevant. Those links are simply the actual URLs that the sitemap will be charged with directing an appliance to.

Content has been removed for optimal crawling.

Below are the links for a given section.

Site Link

<https://extranet.mcplusa.com>

Area Spider Command

</googlesharepointsitemap.aspx?command=site&url=https%3a%2f%2fextranet.mcplusa.com>

Area List Command

</googlesharepointsitemap.aspx?command=lists&url=https%3a%2f%2fextranet.mcplusa.com>

Site Link

<https://extranet.mcplusa.com/departments>

Area Spider Command

</googlesharepointsitemap.aspx?command=site&url=https%3a%2f%2fextranet.mcplusa.com%2fdepartments>

Area List Command

</googlesharepointsitemap.aspx?command=lists&url=https%3a%2f%2fextranet.mcplusa.com%2fdepartments>

Site Link

<https://extranet.mcplusa.com/development>

Area Spider Command

</googlesharepointsitemap.aspx?command=site&url=https%3a%2f%2fextranet.mcplusa.com%2fdevelopment>

Area List Command

</googlesharepointsitemap.aspx?command=lists&url=https%3a%2f%2fextranet.mcplusa.com%2fdevelopment>

If links return HTML errors there will be no way for a search appliance to crawl that content. If this is the case you know that the issue resides either within your GSS configuration.

Server Error in '/' Application.

Runtime Error

Description: An application error occurred on the server. The current cu on the local server machine.

Details: To enable the details of this specific error message to be viewabl should then have its "mode" attribute set to "Off".


```
<!-- Web.Config Configuration File -->

<configuration>
  <system.web>
    <customErrors mode="Off"/>
  </system.web>
</configuration>
```

The next step would be to make sure that all of the proper information has been entered into the **Settings** section of the GSS

The screenshot shows the MC+A web interface. At the top, the logo 'MC+A' is followed by the tagline 'CONNECTING BUSINESS INTELLIGENCE'. Below this is a blue horizontal bar. The main heading is 'Google Sharepoint Sitemap - Web Edition'. A second blue horizontal bar is below the heading. On the left side, there is a vertical navigation menu with the following items: [Home](#), [License](#), [Settings](#) (circled in red), [GSS](#), and [About](#). To the right of the menu, there is a list of status messages: 'The System is running version 2.6 and is up.', 'User is logged in.', 'This version is Licensed.', 'Goto [settings](#) to change your license information.', and 'The crawler will crawl the [GSS](#) section for content.'

Make sure that you verify that the account and account password have access to your SharePoint site. To do this, simply navigate to your SharePoint site using the same credentials that your GSS is using. If the credentials are not the issue then you should contact MC+A support for more help.


CONNECTING BUSINESS INTELLIGENCE

Google Sharepoint Sitemap - Web Edition

Home	SiteUrl (seperate by ,):	<input type="text"/>	YourSharePointSite.domain.com	<input type="button" value="Submit Change"/>
License	UserName:	<input type="text"/>	serviceaccount	<input type="button" value="Submit Change"/>
Settings	Password:	<input type="text"/>	password	<input type="button" value="Submit Change"/>
GSS	Domain:	<input type="text"/>	domain	<input type="button" value="Submit Change"/>
About	ListPage:	<input type="text"/>	100	<input type="button" value="Submit Change"/>
	UseLocalCred:	<input type="text"/>	true	<input type="button" value="Submit Change"/>

4.3 Check the appliance's configuration

If the links are functioning properly in your Google Sitemap, but there is still no SharePoint content being crawled, then you should check to which URLs are being crawled. Often time's people make the mistake of entering the URL of only their SharePoint site and not both URLs. Ensure that this is not the case by opening the administrative console of your appliance and going to **Crawl and Index > Crawl URLs**. You should find the URL of your GSS in the top two fields. In addition to those URLs, you should also see the URL of the SharePoint site in the "Follow and Crawl..." field. If you do not see these URLs enter them now.

Google Mini > Crawl and Index > Crawl URLs [Help Center - Log Out](#) [\[Test Center \]](#)

[Home](#)
▼ Crawl and Index
Crawl URLs
[Crawl Schedule](#)
[Crawler Access](#)
[Proxy Servers](#)
[HTTP Headers](#)
[Duplicate Hosts](#)
[Document Dates](#)
[Host Load Schedule](#)
[Freshness Tuning](#)
[Collections](#)
[Serving](#)
[Status and Reports](#)
[Administration](#)

Start Crawling from the Following URLs: * (Help)

http://YourGSS.com:7000/

example: http://www.myorganization.mycompany.com/ *required

Follow and Crawl Only URLs with the Following Patterns: * (Help - Test these patterns)

http://YourGSS.com:7000/
 http://YourSharePointSite/

example: mycompany.com/ *required

Now, select **Freshness Tuning** from the options on the left. Enter the URL of your GSS and then select **Recrawl These URL Patterns**. Allow 20-25 minutes for the changes to take effect.

Recrawl these URL Patterns ([Help](#))

Enter URL patterns to be recrawled.

URL Patterns:


http://YourGSS.com/

Then, go to **Status and Reports > Crawl Diagnostics** and check to see if your GSS's URL is listed. If you see the sitemap's URL listed proceed to the next step. Otherwise contact MC+A support.

The screenshot shows the Google Mini interface for Crawl Diagnostics. The page title is "Google Mini > Status and Reports > Crawl Diagnostics". The main content area includes a "Specify the URLs for which you want diagnostics:" section with a "Show URLs" button and radio buttons for "Include" (selected) and "Exclude". Below this is a table titled "All hosts" with the following data:

Host Name	Crawled URLs	Retrieval Errors	Excluded URLs
YourGSS.com	435	0	0

Now, run a search for content that would be found in your SharePoint site. If the appropriate search results are not returned contact MC+A support for further assistance.

[Web](#) [Images](#) [Groups](#) [News](#) [Local](#) **Appliance**

[Advanced Search](#)
[Search Tips](#)

Search: public content public and secure content

Powered by Google Search Appliance

Appendix A: About MC+A

MC+A is a technology firm dedicated to solving business challenges by enabling organizations to derive maximum value from their connected information technology assets. With expertise in areas of portal development, enterprise search, software architecture, and supporting technologies – MC+A builds applications, defines processes and leads implementations which help our clients leverage processes and information as a strategic business asset.

A commitment to maximize client satisfaction runs through the core of MC+A; our experienced and highly competent consultants ensure that our work is of the highest quality. With our expertise in technology strategy, systems implementation, software architecture and development and our knowledge of industry trends, best practices and packaged software, we ensure delivery of solutions which are tightly bound to your strategic business objectives.

