

GSA integration manual.

Product version:	1.0
Document version:	1.0
Document creation date:	02-06-2006

Purpose

The GSA integration module uses GSA (Google Search Appliance) as an external search engine. The GSA is setup to crawl and index the content in EPiServer. The search will be made against the GSA and presented within EPiServer. The solution is built upon the Search API module functionality.

Revision Information

Document version	Comments
1.0	Initial version

Table of Contents

INTRODUCTION	3
SYSTEM REQUIREMENTS	3
EPISEVER VERSION.....	3
INSTALLATION	3
CONFIGURATION	3
EPISEVER CONFIGURATION	3
SEARCH API CONFIGURATION	3
GSA CONFIGURATION	4
QUICKSEARCH CONFIGURATION	4

Copyright © ElektroPost Stockholm AB. ElektroPost and EPiServer® are registered trademarks of ElektroPost Stockholm AB. Other product and company names mentioned in this document may be the trademarks of their respective owners.

The document may be freely distributed in its entirety, either digitally or in printed format, to all EPiServer users. Changes to the content or partial copying of the content may not be carried out without permission from ElektroPost Stockholm AB:

ElektroPost Stockholm AB
Finlandsgatan 38
SE-164 74 Kista
Sweden

Changes are periodically made to the document and these will be published in new editions of the document. ElektroPost reserves the right to improve or change the products or programs included in this document at any time.

Introduction

The purpose of the GSA integration module is to provide an example of how GSA search can be integrated in EPiServer. The module uses the Search API to add metadata about access rights (ACL) and page hierarchy to each page when it's crawled by GSA. Then during search this additional metadata is used so only pages which the user has read rights to is served in result set. The metadata about page hierarchy is used to extend quicksearch to be able to search only in a part of hierarchy. It is possible to use the Search API to add more custom meta data if that is needed.

System Requirements

EPiServer Version

- EPiServer 4.51 or later.
- Search API module

GSA Version

- 4.60 or later

Installation

Unpack GSA Search to the <site root> directory of your EPiServer installation. Import the file <site root>\GoogleSearch\GSA_SearchPageType.epi4 to your EPiServer site. Create an instance of page type "GSASearch", which will be your search page.

Configuration

EPiServer configuration

Create a user (or use an existing user) that the GSA crawler will be setup to use when crawling the content. When this users request pages additional metadata will be supplied with the page.

The <site root >\templates\units\Headers.ascx file (or similar if you don't use the standard template package) should be extended with a directive `<%@ Register TagPrefix="development" TagName='MetaData' src='~/SearchHelperService/UserControl/MetaData.ascx' %>` and a tag `<development:MetaData ID="MetaData" runat=server/>` at the bottom of the file. This is so additional meta data can be added to the page.

Search API configuration

Install Search API module if it is not already installed. Create a new connector for GSA. On "General" tab set name to what you like, plugin dll to GoogleSearch and enable it. On tabs "Push settings" and "EPiServer user"

you don't have to specify anything since the GSA will crawl content. On tab "Crawl settings" you set the user to the user you created in the previous section. Then you point out the metadata to be transformed by <site root>\SearchHelperService\Xsl\metatransform.xsl .This xsl will add metadata about ACL and which part of the site the page belongs to. You can change this xsl to your own custom xsl to get other meta data if you like. After you have saved the settings, you will notice that a new tab "Google manager" appears, on this tab you specify which url the GSA search page has (you can optionally also state which collection and frontend to use at search).

After you have made all settings you can control that the metadata is added to the page as it should by logging on to the site as the user that GSA crawler will use. If you view the rendered html source you should see some additional meta data tags in the head section about acl and page part.

GSA configuration

Setup GSA to crawl the EPiServer site. Use "Cookie Site" settings to setup the crawler to crawl as the above given user. When setting up "Cookie Sites" point out <site>/GoogleSearch/templates/Login.aspx as login page (or if your site is running on asp.net 2.0 Login2.aspx) This is because GSA will not handle a Login page that redirects to another page.

QuickSearch configuration

If you want quicksearch to use Google search set dynamic property MainSearchPage to point to the search page created under installation and change the Src attribute for the quicksearch registration in file <site root>/templates/Frameworks/DefaultFramework.ascx to "~/GoogleSearch/templates/Units/QuickSearch.ascx".

Search in included documents

The security model in the integration is designed so that when a page is crawled additional metadata about access rights are appended to the page (and at searches the result is filtered upon this metadata). For linked documents instead of appending metadata to the document itself, the metadata is pushed to the GSA separately. This is done through the scheduled job "Send File Metadata to GSA". It will loop through all documents and push information about access rights to the GSA. Then at search the GSA will only return documents with matching access rights.

It is possible to configure which document types that will be included in the push. This is done on field "Excludes document type" on tab "Push Settings" on Search Engine manager".