DNN Site Search

User Guide

Table of contents

Introduction

The psSiteSearch module is a search engine module for your DNN website. It is based on Lucene.net and includes a web spider which is used to crawl your site and gather all page URLs. This means you can have a real search engine on your own website and not have to rely on your modules providing a search interface to index your site content. This gives you much more control and provides far more accurate search results.

The search engine will honour your robots.txt file and also any page META directives including "no index" and "no follow".

Features

- True web spider crawling and indexing of your site and content.

- Faceted search results by document type and site.

- Search results document type icons to show the user what a search result link points to.

- Honours your robots.txt rules.

- Honours page META directives including "no index" and "no follow".

- Skin object so you can include the search control directly into your skin templates.

- Easy administration with just one module installed and reused.

- Automatic spider schedule using the DNN scheduler.

- Search input box allows the user to enter a search term and just press the enter key to perform the search. Saves users have to click the search icon.

- Option to select a DNN Role for the spider to impersonate when indexing content. Useful for member only areas.

- Option to specify a Windows account for the spider to login with for sites which use Windows authentication.

- Ignore regions

- Link filters to allow the filtering of any URLs which need to be excluded from the index

- Azure compatible

System Requirements

DNN Site Search is compatible with DNN 4.x (v4.0.1 upwards) and 5.x. It requires SQL 2005 and upwards including SQL Azure. You will also need to ensure your DNN portal is running with **Full Trust**. In general, most hosting providers will already set it to Full trust but if you experience problems with the spider creating your search indexes then it is usually necessary to update it via

your web.config file. If you take a look within your web.config file you will see an entry similar to the one shown below:

```
<trust level="Full" originUrl=".*" />
```

You need to ensure that the "level" is set to Full. This is usually set within the machine.config file which your hosting provider will have control over. It is always recommended that you ask your hosting provider to make the change themselves before updating your site web.config and installing this module.

Installation

Install the DNN Site Search module as a normal DNN module. Step by step instructions are shown below.

1. Login in as Host.

2. Go to Host->Module Definitions.

3. Select "Install New Module" from the drop down menu.

4. Select the "psSiteSearch_01.00.00_Install.zip" file and click "Install New Module".

5. If you install the module in DNN 4.x , you can see installation message. If you install the module in DNN 5.x, please follow the installation wizard. If you encounter any errors with the install, please contact technical support supplying the DNN version, SQL version and the error text from the installation page, support@purednnmodules.com

How to use the module

Once you have successfully installed the module, you will see the new module in the "Module" drop-down within the DNN control panel at the top of the page. The Site Search module is listed as "psSiteSearch"

 In order to add the module to the desired page, please follow these steps:

1. Select the "Add New Module" radio button

2. Choose "psSiteSearch" from the "Module" pull down menu

3. Click on the "Add" link (or the arrow icon) to add an instance of the module to the page

Licensing

To be able to use the module you must either apply an Evaluation key or a Purchased key to unlock the module and make it work.

Evaluation license:

To retrieve a 30-day Evaluation license for the module, please register and login to our website:

https://www.puresystems.co.uk/registration

Once logged in, please go to:

https://www.puresystems.co.uk/my-account/my-new-module-licences

From here you can create evaluation licenses. You will also need to add your domain to the licence and update it so that it will run on your evaluation environment.

You will receive an email containing your Serial number. To apply the serial, go to the module settings for the module and enter the serial within the "Serial" field and click the "Validate Serial" button retrieve the license details. Click Update to save the settings.

Purchase a license:

To purchase a license, please follow this link to the DNN Store and purchase the module:

http://store.dnnsoftware.com/home/product-details/dnn-search-engine-pssitesearch/r/c21c511a4de2468cb833

You will receive an email containing your Serial number. To apply the serial, go to the module settings for the module and enter the serial within the "Serial" field and click the "Validate Serial" button retrieve the license details. Click Update to save the settings.

Reassigning or moving the module to a new domain

If you wish to move the licence to a new website then follow these instructions:

1. Login to our website at https://www.puresystems.co.uk/login
2. Go to "My New Module Licences" page, https://www.puresystems.co.uk/my-account/my-new-module-licences
3. Edit your licence in the list by clicking the Edit button
4. In the Domains list below the licence details, delete the old domain and add a new one
5. Click the "Update Licence" button to save your changes
6. Go back to your module within your website and enter the Module Settings
7. Click the "Validate Serial" button which will refresh the licence changes

Quick setup instructions

Please follow the steps below to successfully setup the module

1. Create a new page and call it "Site Search Admin" and make sure that only **Administrators** have access to this page.

2. Add an instance of the psSiteSearch module to this page and enter the module Settings page be selecting "Settings" from the module action drop-down menu or clicking the settings icon below the module.

3. On the module settings page, expand the "psSiteSearch Settings" section and from the "Module View" option, select "Admin" and click the "Update" link to save the settings.

4. You will now see the psSiteSearch admin interface which requires you to add a new site. Click the "Add new site" link to display the form for the new site details. Enter the full URL of your site to index and checking the "Enable Spidering" checkbox. The site URL needs to be in the format of: http://www.mysite.com/ or https://www.mysite.com/

5. If you require the spider to index pages or files which require a DNN login, then you can select the appropriate DNN Role for the spider to impersonate when indexing the site. **Please note:** It is important that you have a valid user (other than the Administrator) belonging to the selected Role so that the spider can index content correctly.

6. If your site requires Windows authentication to access content, you can specify a Windows account by entering the domain (if required) along with the username and password. Once you have saved the admin settings, the Windows account password field will be emptied for security reasons. You **do not** need to re-enter the Windows account Password each time you change the admin settings.

7. Once you have entered the URL and enabled spidering, click the "Update" link to save your settings. At this point, the module will notify the spider schedule to start indexing your site.

8. Now you must create the Search Results page. You can either use the DNN standard Search Results page or create your own.

    a. **Using the DNN Search Results page:**
       Select "Pages" or "Tabs" from the Admin menu and within the list of pages there should be a "Search Results" page. Select the page in the list and click the "View Selected Page" icon to view the page.

    b. Delete the DNN Search Results module from this page by selecting "Delete" from the module action menu. This will leave you with an empty page ready for the psSiteSearch module. Go to step 9 below.

    c. **Creating a new Search Results page**
       Add a new page in the usual way to your DNN portal and give it a suitable name. Remember this name as you will need it later. Go to step 9 below.

9. Add an instance of the psSiteSearch module to the page and enter the module Settings page be selecting "Settings" from the module action drop-down menu or clicking the settings icon below the module.

10. On the module settings page, expand the "psSiteSearch Settings" section and from the "Module View" option, select "Search Results" from the "Module View" option and click the "Update" link to save the settings.

11. To add the Search query module to a page, you can either add another instance of the psSiteSearch module to a content pane where the search box should exist or add the search control via your skin templates as a Skin Object. If you add it via a module instance, then from the module settings, expand the "psSiteSearch Settings" section and select "Search

Control" from the "Module View" option. Once selected you can then select the page where the Search Results module exists from the "Results page" drop-down. Once selected, click the "Update" link to save the settings. If you want to add the search control as a Skin Object, then please refer to the section below.

12. Once you have created and added the relevant modules to your pages, the system will create a schedule which runs every 24hrs and indexes your site. The schedule can be viewed from the "Schedule" page from the "Host" menu. Select the "History" link next to the "psSiteSearch Spider" entry to view the activity of the spider and what pages have been indexed.

Other Admin options

There are other settings on the **Site Search Administration** page which are explained below.

Ignore link text

Checking this option tells the spider to not index any text found in links, i.e.

```
<a href="mypage.aspx">This is the link text and will not be
indexed</a>
```

Debug

Checking this option will create a "log_spider.txt" file in the **/DesktopModules/psSiteSearch/** folder. This is useful when trying to debug any problems the spider is having when indexing your site.

Use File link Titles

With this option enabled, the spider will look for the anchor Titles attribute for any documents found and use this as the Title in the results page if one exists. Default behaviour is to use the document filename.

Index link filters

If you want to exclude certain URLs from being indexed then you can add part of that URL as a link filter so that the spider will ignore any URLs which contain the text. Simply enter the text you want to filter links by into the **Link filter text** text box, i.e.

```
/ctl/
```

and click the **Add** button. It will appear in the list below. You can then edit or delete the entries via the grid controls.

Directories to Index

It is possible to add Portal directories to the content to be indexed by the spider. Simply select the Portal directory from the "Portal Directories" drop-down and click the "Add" button. The selection will appear in the list of added directories below. If you wish to remove a directory from the list, simply click the directory entry and click the delete icon to the right of the list.

The spider will index the Directories added and will also index any child directories so it will **recursively index**. This means you do not need to add child directories if you have already added the parent directory to the list.

Once you have added your portal directories to be indexed, it is important that you select the correct Portal Alias from the "Portal Aliases" drop-down below the list of directories. The portal alias URL is used so that the spider can browse to the directory content as you would via a web browser. The URL is also used to generate the URL to the files in the search results.

Canonical Links

Canonical links tell search engines that the preferred location of this URL is somewhere else. This helps search engines determine which page is the original. The psSiteSearch spider can look for these links and if found it will ignore the current page so as to reduce the chances of duplicate content being indexed.

Enter the regular expression you want to use to find the Canonical Link within the web page. An example of a suitable RegEx is:

```
<link rel="canonical" href="(.*)".*/>
```

If the RegEx field is empty, this means the spider will not look for the canonical link and is not actioned.

iFilter info

By clicking this link the area will expand to show you which iFilters are installed on your web server. This is useful to determine if you have the correct iFilters installed. Some older Microsoft Office documents require the **Microsoft Office Filter** to be installed so that these documents can be indexed.

Ignore Regions

It is possible to mark regions of content in your skins or via a module editor to not be indexed. This is achieved by using ignore region tags, i.e.

```
<!--PSNOINDEX-->
<p>This is the content which will be ignored by the psSiteSearch
spider</p>
<a href="page.aspx">This link will be followed but the text will be
ignored</a>
<!--/PSNOINDEX-->
```

Any content found in the region will be ignored and not indexed, although any links found will be followed. This is useful if you want the spider to ignore any text in your menu system so that you do not get search results appearing for menu text.

Search Results settings

Filter by Role:

Select this option to make the search results only show results which the current user can access.

Sites to Search

Select the websites which you want to display results for in this instance of the search results. This list will show all sites you've added for the current Portal. To select multiple sites, hold down the "Ctrl" key and click the sites you wish to show results from.

Show Descriptions

Option to show the text which the spider has indexed around the keywords/phrases which a user searches for.

Show URL

Option to show or hide the URLs of the results.

URL display length

You can set the maximum length of the URL when displayed in the search results.

Results per page

You can enter the amount of results to show per page.

Max Results

You can enter the maximum number of search results which the search engine will return per query. This is useful if you have a large amount of results being returned but you want to reduce the amount shown to the user. This can also increase performance when set to a lower amount.

Query Operator

This is the Query Operator which the search engine will use when a user performs a search. The OR operator will attempt to find 1 or more of the words entered by the user. The AND operator will attempt to find results where ALL the words entered by the user are found.

For more info, please see:
http://lucene.apache.org/java/2_9_2/queryparsersyntax.html#Boolean operators

Wildcard Search option

If you require the search results to find results using a Wildcard then check the "Wildcard search:" option within the Search Results module settings area. This will append an asterisk to the end of the search query when submitted to the search process. For example, if the user entered "slid" as the search query, the results would return pages which contained "slid", "slide" and "slider". This is a useful option to make the search results more flexible.

Open files in new window

Enabling this opens forces any links to files in the search results to open in a new window by using the `target="_blank"` link attribute.

Show link icons

This option will show icons next to each search results entry denoting the type of results it is, i.e. Webpage, PDF document, Word document, etc.

Faceted Search

This option will enable the search results to display the different document types which have been found from the search query. i.e. PDF Documents, Word Document, Excel Documents, etc. A document count will also be displayed in brackets after the result type so the user can see how the search results are made up. The user can also click the document type and the search results will refresh to only show the selected document type in the results. This allows the user to filter the results by the document type.

Search Control Skin Object

You can add the psSiteSearch skin object to your own skin templates. To do this, follow the instructions below:

1. Open up your skin (.ascx) file you wish to add the Search Control to and add the following code to the top of the page below the first line:

```
<%@ Register TagPrefix="ps" TagName="SITESEARCH"
Src="~/DesktopModules/psSiteSearch/Search.ascx" %>
```

2. You now have to add the actual code to your skin template page where you want the search control to appear :

```
<ps:SITESEARCH runat="server" id="psSiteSearch"
ResultsPageName="Search Results" UseCustomCSS="false"
SearchBoxText="Search..." />
```

The attributes within the skin object code above are explained below:

ResultsPageName
Enter the name of the page within your site which has the psSiteSearch module on it which is set to the "Search Results" module view. This tells the search input box where to send the user once they click search. This will be either "Search Results" if you reused the standard DNN search results page or the name of the new page you created.

UseCustomCSS
You can either let the module style the input box or you can override these styles within your own

skin CSS. If this attribute is set to "false", then the module will add inline styles to the search input box. If this is set to "true", then no inline styles are added and the box is left un-styled. The CSS classes and styles used by the search control are as follows:

```
.psTxtSearch
```
Used to style the search input box

```
.psSearchTxtBtn
```
Used to style the search image button

SearchBoxText (optional)
You can specify text to appear within the search control input box prompting the user to enter a search term. When the user clicks into the input box, the text will disappear. When they leave the input box the text will reappear.

SubmitBtnImage (optional)
You are able to specify the path to the Submit Image if desired. This is particularly useful when you have multiple portals requiring a different button style. An example of the usage is:

```
SubmitBtnImage="~/Portals/0/images/search-btn.gif"
```

Always start the path with a tilde so that the module can find the root of the portal.

RedirectMode (optional)
If you require the search control to send queries to a page outside of the current portal then you can specify that the control should redirect to a URL, i.e.

```
RedirectMode="URL"
```

You then have to update the `ResultsPageName` attribute with the complete URL to the search results page.

Search Control button image

If you wish to replace the search control button image then you can upload your new image into the following location ensuring it is called the same as the existing image below:

\DesktopModules\psSiteSearch\images\search-btn.gif

Or if you are using the Skin Object then it's easier to use the **SubmitBtnImage** attribute above so that any upgrades of the module do not overwrite your image.

Support

If you require support for this module then please contact either by email directly: support@puresystems.co.uk or via the Snowcovered.com helpdesk.

Licences

You are licensed to install this product on one DNN instance only. Additional portals and DNN instances require additional licences. Please see the "license-agreement.txt" document which came with this package before installing and using this software.

This product contains and/or uses third party components.

Lucene.net

Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

DotNetNuke components are included / used under the DotNetNuke BSD licence

Module Version History

Version 02.00.17
Added option to specify the search Submit Button image URL via the Skin object.

Version 02.00.16
Update search results code to handle situation where a home page hasn't been selected in the DNN Site Settings page.

Version 02.00.15
Added ability to direct search queries to a URL as well as a DNN page. Updated Search Results page settings to allow the searching of sites from other portals.

Version 02.00.14
Updated Robots.txt parsing routine. Updated CSS for search input button.

Version 02.00.13
Updated jQuery include routine so it's not unnecessarily loaded on the Search Admin page.

Version 02.00.12
Updated module to use Lucene.Net v2.9.4

Version 02.00.11
Make Azure compatible. Added option to hide URLs in search results. Added option to make search result titles for documents use the anchor title attribute.

Version 02.00.09
Updated query parsing routine.

Version 02.00.08
Updated robots.txt file processing.

Version 02.00.07
Updated DNN6 compatibility for linkclick URLs.

Version 02.00.05
Repackaged installer.

Version 02.00.04
Updated Windows login settings to correctly remember password without the need to re-enter it each time the site settings are updated.

Version 02.00.02
Updated installer scripts. Added Required validator to Search Results module settings for Sites selection.

Version 02.00.00
Updated Search Admin page to allow for multiple sites (DNN and non-DNN) to be added and indexed. Added ability to show search results Facets and filter results by a selected facet. Ability to

show the search results link document type icon. Updated Search Results settings to allow ability to select multiple sites to search. Added Search Results option to open documents in a new window.

Version 01.02.00
Added the ability to filter search results based on the user's permissions.

Version 01.01.21
Added option to add Portal file system paths to the search scope.

Version 01.01.14
Updated module to allow URLs with port number present.

Version 01.01.13
Updated module to allow for multiple child portals.

Version 01.01.11
Updated Lucene indexing procedure.

Version 01.01.10
Updated search results layout to hide paging controls when max results per page is set to 0.

Version 01.01.09
Updated CSS classes for search results presentation.

Version 01.01.07
Updated PDF indexing routine if PDF iFilter isn't available. Faster indexing is achieved of PDFs in this instance and is compatible with 64bit platforms.

Version 01.01.06
Updated spider to allow ignore regions and link text filters to enable the exclusion of URLs containing matched strings.

Version 01.01.05
Updated spider link routine for better parsing of files.

Version 01.01.04
Updated iFilter plug-in to better support PDF and Office documents.

Version 01.01.03
Updated logging options. Updated search results options to specify URL length.

Version 01.01.02
Updated spider agent.

Version 01.01.01
Updated spider to use default document setting to remove duplicate results for the home page.

Version 01.01.00

Updated spider to login using Windows account if required. Added ability to specify DNN Role for spider to impersonate when indexing the site.

Version 01.00.02

Added ability to specify a DNN Role and for the spider to impersonate when indexing the site. Added ability to specify a Windows Domain and User account for sites protected with Windows authentication.

Version 01.00.01

Updates to the indexing routine to fix problems in rare casing with the index files being locked.

Version 01.00.00

First release.