How to configure a content Source to crawl Exchange 2003 public folders in Moss2007.

By Steve Smith, MVP Microsoft Office SharePoint Server , MCT
Combined Knowledge
www.combined-knowledge.com
steve@combined-knowledge.com

This 'How to' guide will take you through setting up a new content source in SharePoint Server 2007 for crawling an Exchange 2003 Public folder.

My server and client topology for this guide was as follows

Server / Client Name	Role	Products Installed
Rootdc	Domain Controller DNS Server Exchange Server	Windows server 2003 SP1 Exchange Server 2003 SP1
SQL	Database server	SQL 2000 SP4
MOSS2007	MOSS2007 Server	MOSS2007 Enterprise
XPCLIENT	CLIENT PC	XP SP2 Office 2007 Enterprise

My internal domain name configured for the Active Directory is trainsbydave.com
You will need to replace trainsbydave.com with your own testing domain name

Step 1 – Let's get the Public folder configured for the correct security

In my testing a decided to create a new public folder called fun. Then in Exchange 2003 System Manager (or you can use outlook) you need to correctly configure the permissions for access to the public folder. It is important that the crawler service has got reviewer rights as a minimum to the public folder plus the people who you wish to search for the document in SharePoint also has a minimum of reviewer rights in order to see the results.

I will demonstrate in this whitepaper how Wilma will have no access to the public folder and hence no search results and Steve will have reviewer rights and hence will get search results.
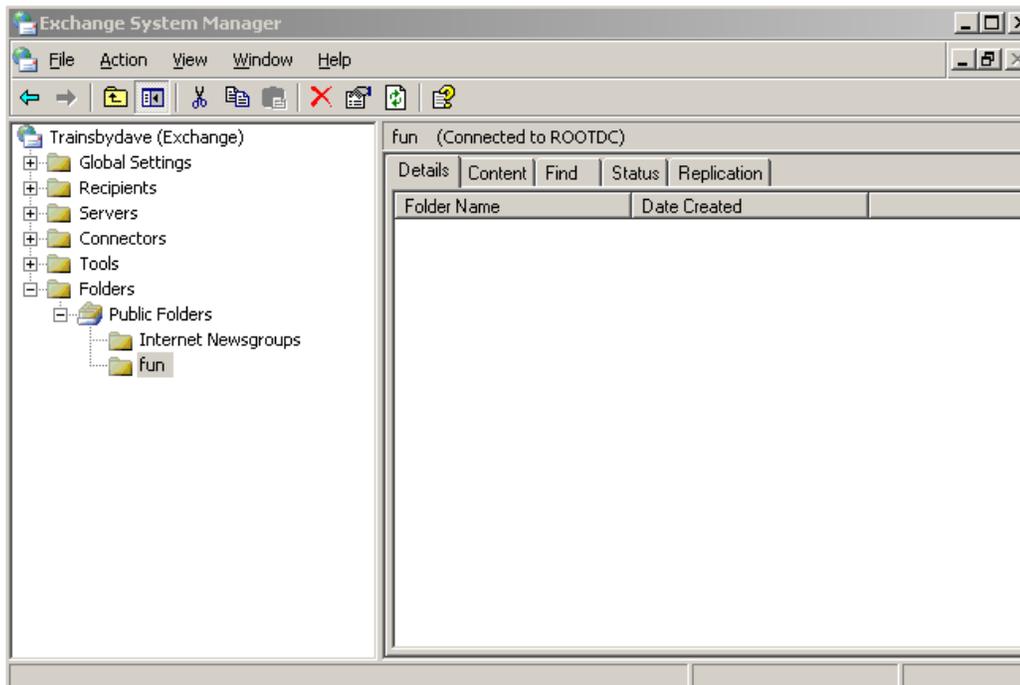
Figure 1.1 – finding the public folder in Exchange System Manager

To correctly configure the permissions follow these steps:

Right Click on the Public Folder name
Click Properties
Select the Permissions tab
Now add the users / groups of the people to allow access to the Public folder.
Change the default and anonymous to none unless you are happy for everyone to see the search results.
Ensure that the crawl service account has reviewer access to the folder or nothing will be indexed

In order to achieve my results I gave Wilma None and Steve Reviewer rights. See figure 1.2.
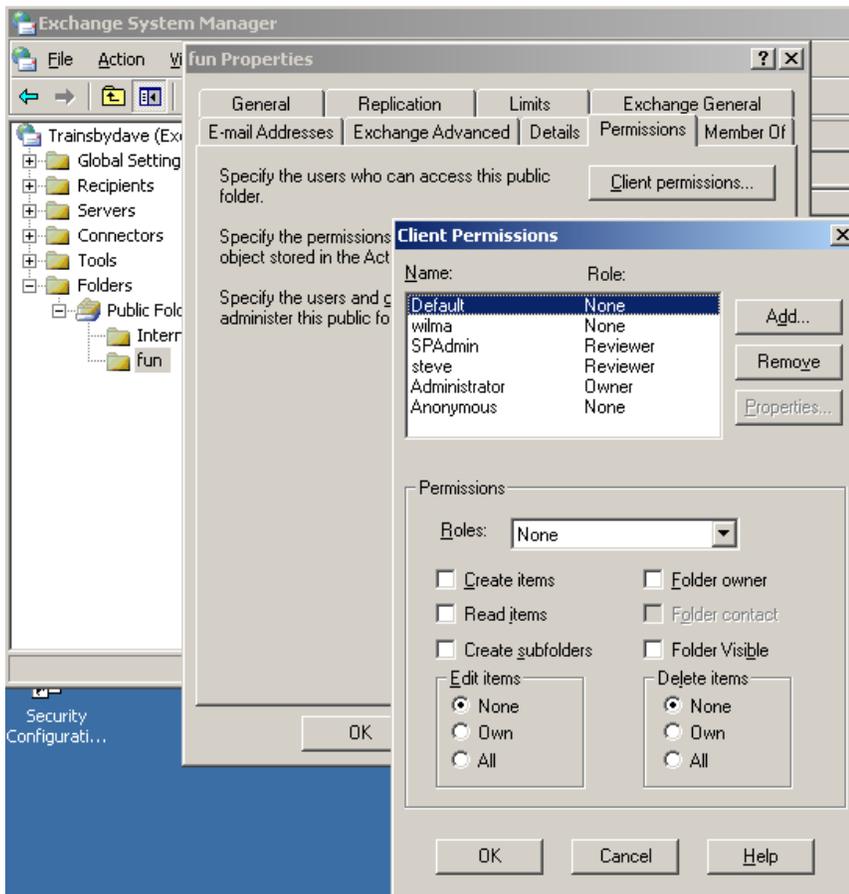
Figure 1.2 – Assigning rights to the public folder

Now that the public folder permissions are set we need to now configure a new content Source in sharepoint Central Administration.

So what is a content source some of you may ask?

In order for SharePoint to go off and index (crawl) documents in a source location other than SharePoint the Shared Services Provider (SSP) administrator needs to go the SSP Admin site and create pointers to the location you wish to crawl. This can include Public folders in Exchange – http://exchange/public , File shares file://server3/share , Web Sites http://combined-knowledge.com and other sources connected by the business data catalog such as an HR database or SQL data source such as a products database.

In order for a content source to crawl correctly there are several items that need to be in place.

1. A crawl account that has read access to the source.
2. A I Filter that is able to open the document and then index it, such as the Adobe PDF I Filter.
3. A Protocol Handler that is able to actually communicate with the source. You will notice that for Exchange Public Folders we use http and not Mapi as the protocol handling method.

For this whitepaper we need to create an Exchange Public folder content source.

Follow these steps to create the Content source

1. Go to **Central Administration**
2. Go to the **SSP Administration page** – Do this by clicking on the name of the Shared Service Provider you want to configure the source on from the left hand quick launch link on the home page of central administration or go to Application management page and click on 'manage this farms shared services' and click on the name of the SSP to open the SSP administration site.
3. In the SSP Administration page click on **Search Settings**
4. Click on **Content Sources**
5. Click **New Content Source**
6. There are now several fields that need to be completed:
   a. Name – Use a good descriptive name for the content source (you will need it later)
   b. Content Source type – We will select Exchange Public Folders
   c. Start Addresses – we now need to tell this content source what to go and index. You must use the format http://exchangeservername/publicfoldername in order to crawl the content. You can however multiple sources here in order for this crawl rule to index multiple public folders in one schedule.

> ****Note**** you must remember that you can schedule content sources but not individual locations in the content source. Therefore do not include say 20 public folders if they all take 8 hours to index. Consider creating multiple content sources with different locations and different schedules in order to break down the indexing matrix so that other jobs can run in the overnight timeframe. Also consider if the content in the public folder changes as some sources may be static and therefore only requires a one of full crawl and no schedule. This would be a good example of why you would want to create separate content sources.

7. Crawl settings - Here you can define if you want to index just the content at the URL you specified of if you want to also crawl all sub folders. This is very useful if you have a root public folder with 100 sub folders but only want to index 10 of the sub public folders so you can create one content source with each of the 10 folders listed in the start address field and set the crawl to 'The folder of this start address only'
8. Crawl schedules – define your full and incremental crawl schedules if you wish to index the content at regular intervals.
9. Start Full crawl – this checkbox will initiate a full crawl of the define start addresses
10. When happy Click ok

See figure 1.3 on the next page

Add Content Source

Use this page to add a content source.

* Indicates a required field

Name

Type a name to describe this content source.

Name: *

ExchangePF

Content Source Type

Select what type of content will be crawled.

Note: This cannot be changed after this content source is created since other settings depend on it.

Select the type of content to be crawled:

○ SharePoint Sites
○ Web Sites
○ File Shares
● Exchange Public Folders
○ Business Data

Start Addresses

Type the URLs from which the search system should start crawling.

This includes messages, discussions and collaborative content.

Type start addresses below (one per line): *

http://rootdc/public/fun

Example:
http://exchangeserver/public/folder/subfolder

Crawl Settings

Specify the behavior for crawling this type of content.

Choose which folders to include in the crawl.

Select crawling behavior for all start addresses in this content source:

● The folder and all subfolders of each start address
○ The folder of each start address only

Crawl Schedules

Select the crawl schedules for this content source.

Select the schedule that this should be a part of:

Full Crawl

At 00:00 every Sat of every week, starting 28/06,▼
Edit schedule

Incremental Crawl

At 02:00 every day, starting 28/06/2007 ▼
Edit schedule

Start Full Crawl

Select "Start full crawl of this content source" and click "OK" to start a full crawl of this content source.

☑ Start full crawl of this content source

OK   Cancel

Figure 1.3 Creating a Content source

Now we need to manage the content source and check the crawl log to make sure that the documents in the public folder have been indexed correctly. Follow these steps in order to manage the content source. See Figure 1.4

1. In search settings click content Sources
2. Mouse over the exchange Public Folder and click on the drop down arrow to the right
3. You now have several options for managing the content source.
    a. Edit – change the settings such as adding addresses, changing the schedule etc. as per the new content source page.
    b. View crawl Log – see the results of all items indexed
    c. Start full crawl – manually force a full crawl
    d. Start Incremental crawl – manually run a crawl for changed and deleted content only
    e. Resume / Pause / stop crawl – allow you to manage the crawls in real time.
    f. Delete – deletes the content source.

Shared Services Administration: CorpSSP1 > Search Settings > Content Sources

Manage Content Sources

Use this page to add, edit, or delete content sources, and to manage crawls.

New Content Source | Refresh

Type	Name	Status	Next Full Crawl	Next Incremental Crawl
	Local Office SharePoint Server sites	Idle	None	None
	filearchive	Idle	None	None
	exchangePF	Idle	None	None

Edit
View Crawl Log
Start Full Crawl
Start Incremental Crawl
Resume Crawl
Pause Crawl
Stop Crawl
Delete

Figure 1.4 – options for the content source and launching the crawl log

Let's now have a look at the crawl log to ensure we indexed the content.

1. Click on View Crawl log
2. The hostname summery > URL Summary page should now be displayed.
3. Notice that by default status type is set to all, in other words we are seeing every item indexed returned in this view. You can however filter this view by selecting options from the drop down arrows.
4. We can see that several docx, xlsx and pptx documents have been indexed and the green icon indicates that the items have been indexed successfully

See Figure 1.5 below

Figure 1.5 – viewing the crawl log

Ok so we can see the one document called exchangetest has been indexed in the public folder, let's now test the permissions as we set earlier by seeing what Steve and Wilma see when doing a search.

Logon as Wilma to the home page and run a search for exchangetest (in my test exchangetest is the name of a document in the public folder).
You will notice as per Figure 1.6 that it returned no results. This is because Wilma has No rights to the Public Folder. It is also important to remember that anonymous and default also has no rights to the folder.
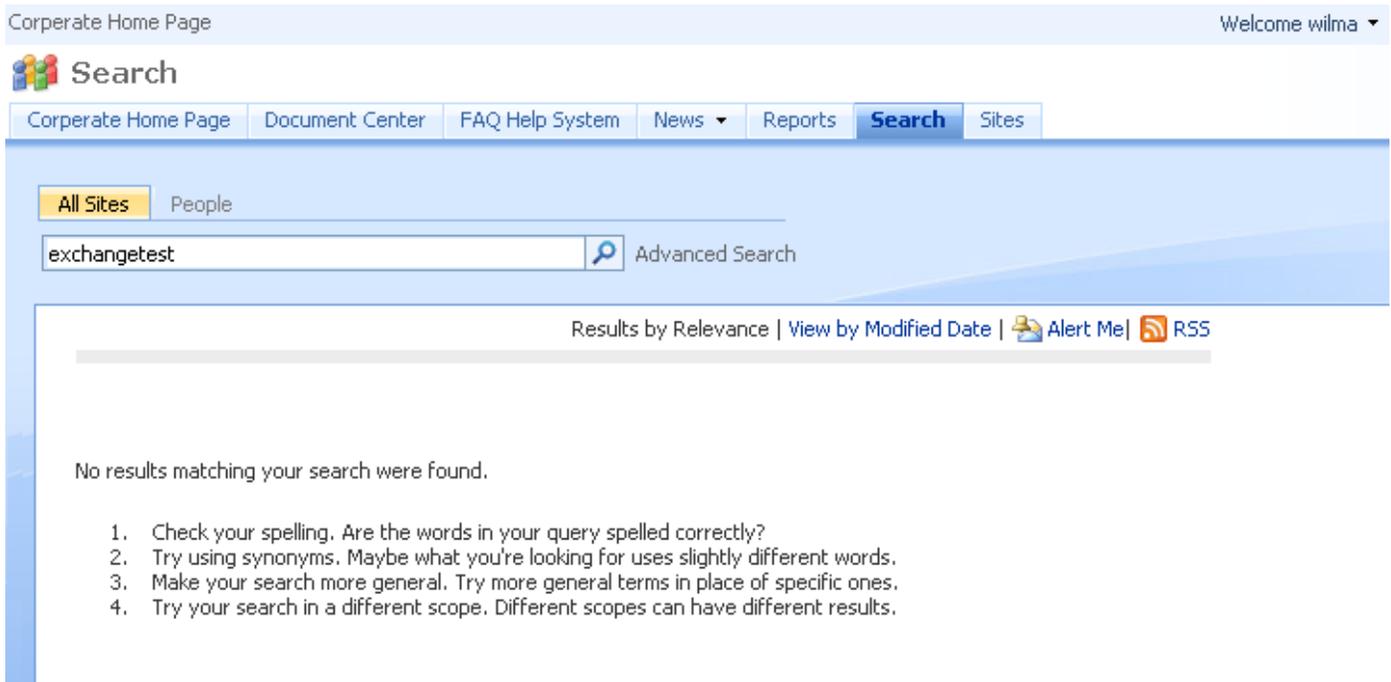
Figure 1.6 - Wilma getting zero search results.

Now let's do the same test but logged on as Steve. Notice as per figure 1.7 Steve gets to see the returned document. This is because Steve had reviewer rights on the public folder thus proving that the security trimming is working through SharePoint Search to the indexed Exchange Public Folder.

To further enhance a user's search experience you could also go on to create a Search Scope that the users could select so that a search was only retrieving results from our content sources pointing at Exchange public Folders. I will cover this in another whitepaper so keep an eye out for that soon.
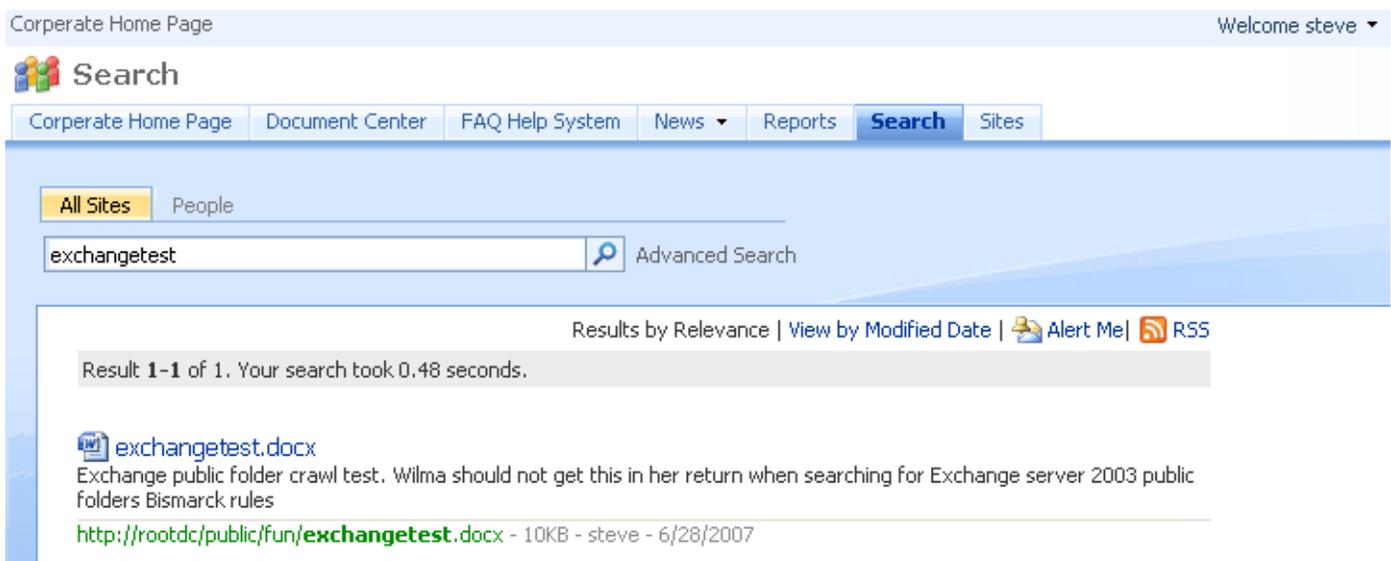
Figure 1.7 - Steve running the same search but this time getting a result

I hope you found this article useful , please do feel free to send me comments and keep checking our web site for more downloads on configuring some of the cool new features of Moss2007.

Steve@combined-knowledge.com