

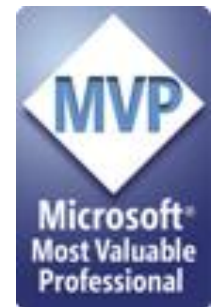


## *Stuff about Search*

Shane Young, MVP  
Chief Storytelling Officer  
SharePoint911

# Who Am I?

- Shane Young
- Owner of SharePoint911.com
- Microsoft Office SharePoint Server MVP
- Consultant, Trainer, Writer, & Speaker
  - [shane@sharepoint911.com](mailto:shane@sharepoint911.com)
  - **Blog**
    - <http://msmvps.com/shane>
  - **SharePoint Consulting**
    - <http://www.sharepoint911.com>
  - **Social Networking**
    - Facebook, LinkedIn, Twitter @shanescows



# Agenda

- WSS vs. MOSS
- OOTB Search
- Administrate Search
- Customize Search Options

# MOSS Search is Awesome!

- WSS Search is cool too!
- Same technology behind both
- WSS Search only works within the site collection
- MOSS can crawl just about anything
- Biggest difference?
  - WSS Search just runs
  - MOSS Search has lots of configuration options
- Remember MOSS Search and SSP

# Search Server 2008

Throw away your search appliances!

Provides all of the lovely MOSS Search features in a free package.

Search Server Express is free

Search Server is purchase but allows high availability

<http://www.microsoft.com/enterprisesearch/serverproducts/searchserver/default.aspx>

## How search works

- Locates content
- Loads protocol handler
- Loads necessary ifilter
- Word breaks the file
- Writes info to index (physical file)
- Writes property info to search db
  
- Search honors security!

## ifilters

- Core Microsoft search technology
- Allows SharePoint to index other file types
- OOTB only indexes
  - Plain text files
  - Word, PowerPoint, Excel files
- Places to find ifilters
  - File Manufacturer
  - 3<sup>rd</sup> party web sites
  - [www.ifilter.org](http://www.ifilter.org) has list of free ones

## Adding a new file type

- Install appropriate ifilter
- Add to SSP > Search Settings > File Types
- Add custom icon file
  - `..\12\template\xml\docicon.xml`
  - Upload to `..\12\template\images (16x16)`
- Double check file type is not excluded for server
  - `Central Admin > Operations > Blocked file types`

## Noise words

- These are words that have no search value
- Files located in two places
  - C:\program files\Microsoft Office Servers\12.0\data\applications\{guid}\config
  - C:\program files\Microsoft Office Servers\12.0\data\config
- WSS change to
  - C:\program files\common files\Microsoft Shared\web server extensions\12\
- Different version for each language
  - Noiseenu.txt for US English
  - Noiseneu.txt for neutral files

# Thesaurus File

- Allows expansion and replacement on queries
- Can also effect search weighting
- Files located in two places
  - C:\program files\Microsoft Office Servers\12.0\data\applications\{guid}\config
  - C:\program files\Microsoft Office Servers\12.0\data\config
- WSS change to
  - C:\program files\common files\Microsoft Shared\web server extensions\12\
- Different version for each language
  - Tsenu.xml for US English
  - Tsneu.xml for all files

## Feature Breakdown - Content Sources

- WSS – Just your SharePoint Site Collection
- MOSS Standard– All of your SharePoint Sites at once and add things like
  - Exchange Public Folders
  - Other Web Sites
  - File Shares
  - Other SharePoint Sites
- MOSS Enterprise – The BDC baby!

## Feature Breakdown - Scopes

- WSS not really – MOSS for sure.
- What is a scope?



All Sites Scope \*

SharePoint Sites, Exchange PF, File Shares, Web Sites, Etc

People Scope \*

All people

HR Scope

<http://portal.tpg.local/hr>

\\server\HR

Exchange PF HR

F: Drive Scope

\\server\share

\* Automatically Created Scopes

## Feature Breakdown - Scheduling

- Indexing is one of the most resource intensive things you can do with SharePoint. Controlling when it happens can be key.
- WSS – Does a full crawl once a day when you choose.
- MOSS – Complete scheduling control. Full and incremental options.

## Other MOSS only features

- Detailed easy to use logs
- Control over relevancy
- Crawl Rules
- Easy to grow out farm and off load roles
- Management of properties and advanced search
- Keywords and best bets
- Search center and 11 search web parts
- Detailed search reporting

# Gather logs

- The place to figure out why search isn't behaving
- SSP > Search Settings > Crawl Logs

Shared Services Administration: SharedServices1 > Search Settings > Hostname Summary > URL Summary

## Crawl Log

Find URLs that begin with the following hostname/path:

Show messages occurring later than:

Show messages occurring earlier than:









Content Sources:

Status Type:

Last Status Message:  [Filter](#)

[Show complete status message](#)

Crawled Content Status - 1227 Items: 882 Successes, 1 Errors, 344 Warnings 1 - 100

URL	Last Content Source	Last Time Crawled
 <a href="http://demo.sharepoint911.com/spa/...efile_movie.htm">http://demo.sharepoint911.com/spa/...efile_movie.htm</a> Crawled	Local Office SharePoint Server sites	10/9/2007 5:02 AM
 <a href="http://demo.sharepoint911.com/spa/...licensefile.htm">http://demo.sharepoint911.com/spa/...licensefile.htm</a> Crawled	Local Office SharePoint Server sites	10/9/2007 5:02 AM
 <a href="http://demo.sharepoint911.com/spa/...ler/en-us/flash">http://demo.sharepoint911.com/spa/...ler/en-us/flash</a> Crawled	Local Office SharePoint Server sites	10/9/2007 5:02 AM
 <a href="http://demo.sharepoint911.com/spa/...er/en-us/images">http://demo.sharepoint911.com/spa/...er/en-us/images</a> Crawled	Local Office SharePoint Server sites	10/9/2007 5:02 AM
 <a href="http://demo.sharepoint911.com/spa/...licensefile.htm">http://demo.sharepoint911.com/spa/...licensefile.htm</a> Crawled	Local Office SharePoint Server sites	10/9/2007 5:02 AM
 <a href="http://demo.sharepoint911.com/spa/...tingstarted.htm">http://demo.sharepoint911.com/spa/...tingstarted.htm</a> Crawled	Local Office SharePoint Server sites	10/9/2007 5:02 AM
 <a href="http://demo.sharepoint911.com/spa/...en-us/about.htm">http://demo.sharepoint911.com/spa/...en-us/about.htm</a> Crawled	Local Office SharePoint Server sites	10/9/2007 5:02 AM
 <a href="http://demo.sharepoint911.com/spa/...s/contactus.htm">http://demo.sharepoint911.com/spa/...s/contactus.htm</a> Crawled	Local Office SharePoint Server sites	10/9/2007 5:02 AM

## MOSS Includes 11 Search Web Parts

- Create custom search solutions
- Complete control over user search experience
- Search center provides starting point
  - **Part of Collaboration Portal Template**

## Keyword/Best bets

- Highly under utilized
- Keyword is what your user is searching for
  - Can assign a definition to display
- Best bet is what you think they want
  - Links to any web url
- Administered by site collection admin

# Advanced Search

- Built with configurable Web Parts

## Advanced Search

### Find documents with...

**All** of these words:

The exact **phrase**:

**Any** of these words:

**None** of these words:

### Narrow the search...

Only the **language(s)**:  French  
 German  
 Japanese  
 Spanish

Result **type**

### Add property restrictions...

Where the Property...     [Add Property...](#)

# Customizing Advanced Search

- Add your own properties
  - SSP > Search Settings > Metadata Property Mappings
- Edit Advanced Search Box WP
  - Properties section is XML
  - Add property definition
    - `<PropertyDef Name="YourProperty" DataType="text" DisplayName="YourPropertyDisplayName"/>`
  - Add property reference
    - `<PropertyRef Name="YourProperty" />`

## Other happy user features

- Hit highlighting
- Did you mean
- Duplicate collapsing

# Administrating Search

- Central Administration
  - Define services/server roles
  - Proxy & Crawler impact rules
- Site Collection
  - Keywords/best bets
  - Local scopes
  - The UI
- SSP
  - Everything else



## Search Reporting

- Available from the SSP
- Two types of reports
  - Search queries
  - Search results
- Great feature for understanding your environment

## Need to customize more?

- Use XSLT and CSS to transform search results
- Write code against Object Model or Web Services
- Look to 3<sup>rd</sup> party
  - <http://www.codeplex.com/FacetedSearch>
    - Free Community Tool
  - <http://www.ontolica.com/>
    - Free Wild Card search/Pay full product
  - <http://www.coveo.com>
    - Replaces SharePoint Search Engine
  - <http://www.ba-insight.net>
    - Hit Highlighting, preview
    -

## The MOSS Search Service

Indexer Performance Setting	Total # of threads	Maximum # of threads per host*
Reduced	# of CPUs	# of CPUs
Partially Reduced	4 times # of CPUs	# of CPUs plus 4
Maximum	16 times # of CPUs	# of CPUs plus 4

\* Can be increased to 64 per host with crawl rule

\*\* 64 is maximum amount of the threads index can have at one time

# Demo!

- Everything search

## Books by the SharePoint911 Consultants

- Real World SharePoint 2007: Indispensable Experiences From 16 MOSS and WSS MVPs (Wrox)
- Professional SharePoint 2007 Design (Wrox)
- Professional Microsoft Search: SharePoint 2007 and Search Server 2008 (Wrox)
- Inside SharePoint Administration (Cengage, Coming soon)
- MOSS Explained: An Information Workers Deep Dive into Microsoft Office SharePoint Server 2007 (Cengage, Coming soon)

<http://www.sharepoint911.com/Links/Pages/Bookshelf.aspx>

# SharePoint911

- SharePoint911 was established in 2004 to draw on Shane Young's extensive experience with systems administration and with consulting for clients on SharePoint issues. Over the past several years, SharePoint911 has expanded and evolved. While our ability to respond quickly to specific needs makes us unique, we also offer comprehensive consulting services. Not only do we have extensive, real world experience, but our consultants also teach, write about, and speak on various SharePoint topics.

*We truly are “The SharePoint Authority.”*

<http://www.sharepoint911.com>