
Aquaforest Searchlight Release Notes

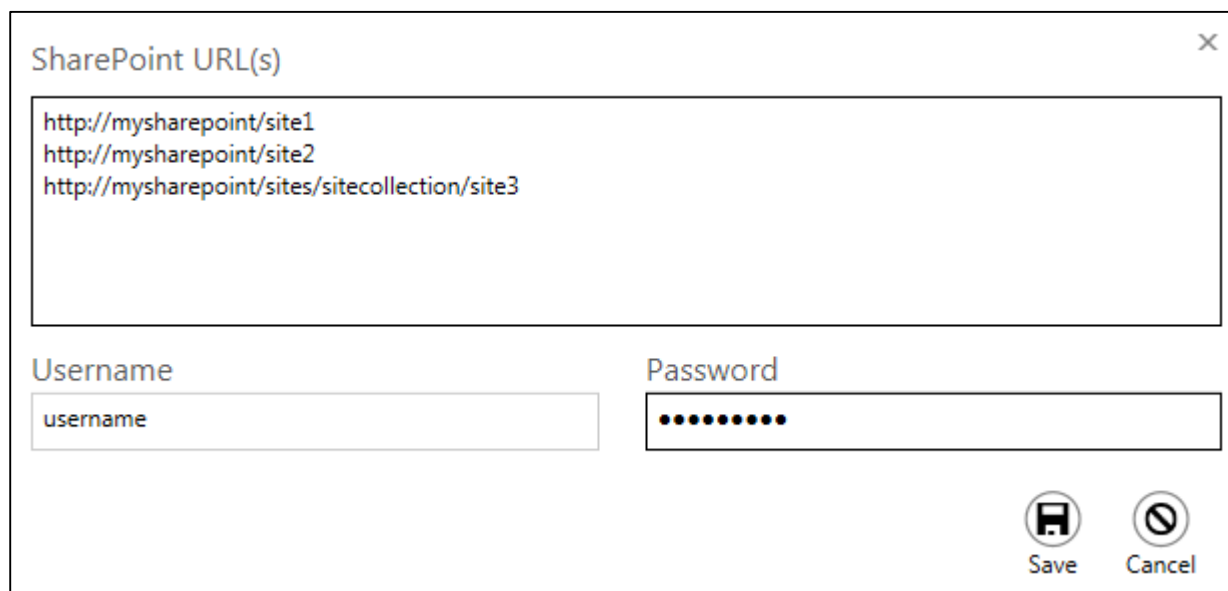


Version 1.05
December 2014

1 Enhancements

1.1 Add Multiple SharePoint URLs

Multiple SharePoint URLs can now be added at once using the new enhanced Add New Location wizard. Each URL must be in a new line as shown below.



SharePoint URL(s)

http://mysharepoint/site1
http://mysharepoint/site2
http://mysharepoint/sites/sitecollection/site3

Username: username

Password:

Save Cancel

1.2 Download Progress

The dashboard now displays the progress when downloading documents in the following format: "Downloading x of y".

1.3 Download Retries

Occasionally, there might be some intermittent network problems which can cause problems when downloading files from SharePoint for processing. To cope with this, a retry mechanism has been implemented that will retry downloading in the event of such network problems. The number of retries and the amount of time to wait between retries can be controlled through the following config setting:

```
<add key="downloadRetries" value="5,1000" />
```

The value needs to be entered in the format "x,y", where x is the number of retries and y is the amount of time in milliseconds to wait for each retry.

This config setting can be found in the "Searchlight.config" file located at: "[installation path]\config\Searchlight.config".

1.4 Database Update Retries

Sometimes, if a document library is set to process using multiple cores, Searchlight may encounter problems when it tries to update the database due to it being 'locked' because of concurrent updates. To overcome this problem, a retry mechanism has been implemented that will retry updating the database if it fails the first time. The number of retries and the amount of time to wait between retries can be controlled through the following config setting:

```
<add key="databaseRetries" value="5,1000" />
```

The value needs to be entered in the format "x,y", where x is the number of retries and y is the amount of time in milliseconds to wait for each retry.

This config setting can be found in the "Searchlight.config" file located at: "[installation path]\config\Searchlight.config".

1.5 Form-based authentication

Searchlight can now process SharePoint libraries that require form-based authentication.

1.6 Remove Hidden Text

Existing hidden text (text that was added as a result of a previous OCR) can now be removed from the PDF file so that the resulting searchable PDF file does not have two layers of the same text. This can be achieved by setting the "Remove Hidden Text" option to True.

1.7 Remove Visible Text

Visible text (text as a result of conversion from an electronic document such as Word to PDF) can now be excluded from the OCR process. This only affects engine 2 of Aquaforest OCR and the Extended OCR (IRIS engine).

To enable this feature:

- Aquaforest OCR - set "PdfToImageIncludeText" to False in properties.xml
- Extended OCR – set "Remove Visible Text" to True from General OCR Settings in the GUI.

1.8 Retain Creation/Modified Date/User

In this release of Aquaforest Searchlight, there is the extended functionality of retaining created date, modified user, created user and modified user of documents.

	Creation Date	Created User	Modified Date	Modified User
SharePoint	✓	✓	✓	✓
PDF metadata	✓	✓	✓	N/A
Windows File System	✓	✓	✓	N/A

"Create User" maps best to "Owner" in Windows File System metadata. For this to be manipulated Searchlight would need to be running with sufficient administrative privileges.

Note: Previous versions of Aquaforest Searchlight had two options "Retain SharePoint TIFF Creation Date" and "Retain Creation Date" which have now been merged to one option namely "Retain Creation Date". If any of the two options were set to 'True' in the previous version, it will be carried over to the new field.

1.9 Multicore support

In this version, the support for multicore processing has been increased from 8 cores to 64 cores.

2 Bug Fixes

2.1 SharePoint Template Types

In previous versions of Searchlight, only document libraries and lists with Server Template IDs 101 and 100 respectively were processed. As a result, document libraries and lists created using custom templates were skipped. This has now been fixed.