
Aquaforest Searchlight Reference Guide



Version 1.05
December 2014

Contents

1	Product Overview	1
1.1	The Business Problem: Documents that are not searchable.	1
1.2	The Solution: Aquaforest Searchlight	1
1.3	Supported Document Stores.....	2
1.4	Supported Operating Systems	2
1.5	Aquaforest Searchlight Modules	2
1.5.1	Multi-Core Module (Included as Standard).....	2
1.5.2	Extended OCR Module (Included as Standard).....	2
1.5.3	Extended OCR Asian Language Module (Extra Cost)	2
1.5.4	Extended OCR Advanced Compression Module (Extra Cost).....	2
1.5.5	SharePoint Metadata Extractor (Extra Cost)	2
1.5.6	Advanced PDF iFilter (Extra Cost)	2
2	Installation and Licensing.....	3
2.1	Installation.....	3
2.2	System Requirements	3
2.2.1	SharePoint Online (Office 365) System Requirements.....	3
2.3	Microsoft SharePoint Requirements.....	3
2.4	Licensing.....	3
2.4.1	Entering License Keys.....	3
2.5	Trial License Mode	4
3	Searchlight Architecture	4
3.1	Processing Model.....	4
3.1.1	Audit and Candidate Identification.....	4
3.1.2	Searchability Status	4
3.1.3	Supported Formats.....	4
3.2	Aquaforest Searchlight Windows Service.....	4
4	Document Library Management.....	5
4.1	Creating a Document Library	5
4.1.1	Select Library Locations	5
4.1.2	SharePoint URL format	6
4.1.3	Document Settings.....	6
4.1.4	OCR Settings.....	7
4.1.5	Scheduler.....	8

4.2	Editing a Document Library	9
4.3	Deleting a Document Library	9
4.4	Processing/Running a Document Library	9
4.5	Review Audits & Conversions	9
4.6	Reports.....	10
4.6.1	Error documents.....	10
4.6.2	Library Status	10
4.6.3	Generating Log files	11
5	Microsoft SharePoint Document Stores.....	12
5.1	Versioning.....	12
5.2	Archiving	12
5.3	Metadata.....	12
5.4	Security.....	12
6	File System Document Stores.....	13
6.1	File Name Length.....	13
6.2	File Access Permissions	13
6.3	Archiving	13
6.4	Metadata.....	13
7	Aquaforest Searchlight Modules.....	14
7.1	Multi-Core Module (Included with the standard product).....	14
7.2	Extended (IRIS) OCR Module (Included with the standard product)	14
7.3	Extended OCR Advanced Compression (Extra Cost Option)	14
7.4	Metadata Extraction Module (Extra Cost Option).....	15
7.5	Advanced iFilter Module (Extra Cost Option).....	15
8	Aquaforest Searchlight User Interface	16
8.1	Getting Started.....	16
8.2	Aquaforest Searchlight Dashboard	17
8.2.1	Dashboard Fields	17
8.3	Library	18
8.3.1	Library Status	18
8.3.2	Library Settings	19
8.3.3	Document Settings.....	20
8.3.4	Aquaforest OCR Settings.....	22
8.3.5	Extended OCR Settings.....	24
8.3.6	Advanced Compression Settings.....	28
8.3.7	Scheduler Settings	29
8.3.8	Alert Settings.....	30

8.3.9	Run Details	30
8.3.10	Run Details Context Menu.....	31
8.4	Support	31
8.5	Diagnostic Tool.....	31
8.6	Settings	32
8.6.1	License Settings.....	32
8.6.2	Email Settings.....	33
8.6.3	Themes	33
9	Aquaforest Searchlight Installation Files.....	34
10	Acknowledgements.....	35

1 Product Overview

Aquaforest Searchlight is an in-place document processing tool that is designed to monitor and make files within an organization Searchable.

It is able to integrate with Microsoft SharePoint, Windows File Systems and File System-based Document Management Systems such as Worldox.

Aquaforest Searchlight revolves around **Document Libraries**. A Document Library in Aquaforest Searchlight is an object that has references to one or more Document Management System locations. Aquaforest Searchlight then monitors these locations and makes sure that the TIFF and PDF documents contained in them are made searchable.

1.1 The Business Problem: Documents that are not searchable.

Studies have shown that in most organizations over 20% of documents are not fully text searchable so will not be located by text search or discovery exercises. In addition a greater percentage of documents may not be tagged with appropriate metadata. With the increase in distributed capture and ad-hoc publishing to document stores such as Microsoft SharePoint, there is a need for a solution to this problem that doesn't require a strict capture-time process.

Many types of documents are not searchable without special processing. For example:

- Scanned TIFF Files
- Image PDF Files
- Faxes

These types of file need to be processed with Optical Character Recognition (OCR) technology to create a text version of the file contents which allows a searchable PDF to be created by merging the original page images with the text. The text is stored in the PDF file as a hidden layer overlaying each page image. This enables the file to be searched.

Documents stored in Microsoft SharePoint may often be lacking key metadata required to enable straightforward metadata searches. For example, attributes such as "Keywords" or "Company" may not have been fully indexed when the document was stored in SharePoint. The Aquaforest Searchlight Metadata Extractor module can be configured to automatically add metadata to new and existing documents.

In order to enable searches across files in SharePoint, Windows Search or other Document Management Systems the searchable files need to be indexed by the system. System iFilters manage this automatically for Microsoft Office but for PDF files a separate iFilter is required. A free iFilter is available from Adobe which does a good job but only indexes basic PDF content, not PDF titles, subjects, authors, keywords, annotations, bookmarks, attachments, create time/date, number of pages.

1.2 The Solution: Aquaforest Searchlight

- *Audits document stores to determine which documents require processing*
- *Document Stores are monitored to deal with new and updated documents.*
- *Dashboard provides a convenient summary of the state of all managed stores.*
- *Provides detailed conversion reporting.*
- *convenient GUI which enables management of all stores via a single interface*
- *OCR Support for 23 languages including English, Spanish, German, French*



1.3 Supported Document Stores

- Windows File Systems
- SharePoint 2010
- SharePoint 2013
- SharePoint Online (Office 365)
- Worldox and other Document Management Systems where files are held in a Windows file system

1.4 Supported Operating Systems

Windows 2008 Server, Windows 2012 Server, Windows 7, Windows 8

1.5 Aquaforest Searchlight Modules

<p>Aquaforest</p>   	<p>1.5.1 Multi-Core Module (Included as Standard) Take full advantage of multi-core architectures for maximum performance.</p> <p>1.5.2 Extended OCR Module (Included as Standard) Adds the IRIS OCR Engine for more comprehensive Language support (over 100 languages) and enhanced accuracy for certain document types.</p> <p>1.5.3 Extended OCR Asian Language Module (Extra Cost) Adds support for Korean, Japanese, Simplified Chinese & Traditional Chinese.</p> <p>1.5.4 Extended OCR Advanced Compression Module (Extra Cost) Adds support for advanced Mixed Raster Compression that can dramatically reduce the size of color scanned documents.</p> <p>1.5.5 SharePoint Metadata Extractor (Extra Cost) Adds the Pingar Metadata Extractor for SharePoint with Aquaforest Searchlight integration.</p> <p>1.5.6 Advanced PDF iFilter (Extra Cost) Adds the Foxit PDF iFilter for maximum indexing performance and PDF searchability.</p>
--	---

2 Installation and Licensing

2.1 Installation

The product is installed via the aquaforest_searchlight.msi install package that can be downloaded from Aquaforest's download page: http://aquaforest.com/en/searchable_pdf_download.asp

2.2 System Requirements

Supported Operating Systems	Windows 7 (32-bit 64-bit) Windows 8 (32-bit 64-bit) Windows Server 2008 R2 (32-bit 64-bit) Windows Server 2012
Disk Space	480 MB
Memory	Minimum 4GB (recommended 8GB)
Visual C++ Redistributable	Visual C++ 2008 Redistributable (x86) and Visual C++ 2012 Redistributable (x86)
.NET Framework	3.5 and 4.5

2.2.1 SharePoint Online (Office 365) System Requirements

Supported Operating Systems	Windows 7 SP1 and above (32 bit 64 bit) Windows 8 (32bit 64bit) Windows Server 2008 R2 SP1 and above (32bit 64bit) Windows Server 2012
Additional tools	SharePoint Server Client Components SDK

2.3 Microsoft SharePoint Requirements

In order to make use of the SharePoint functionality, Microsoft SharePoint 2010 or 2013 (Standard or Enterprise) is required. Note that the metadata service is required before the Entity Extraction module can be used.

2.4 Licensing

2.4.1 Entering License Keys

License keys may be entered by using the "settings" tab from the Aquaforest Searchlight application. Without a purchased license key the product operates in trial mode. The trial license provides access to the following modules:

- Aquaforest OCR
- Extended OCR enabling users to choose the IRIS OCR engine.
- Asian Language Support.
- Multi-Core Module for parallel processing.

Further Modules are also available upon request, these are:

- Intelligent High Quality Compression
- Pingar Integration
- iFilter Integration

2.5 Trial License Mode

In this mode there is a limit of 100 documents that may be converted although there is no limit on the number of documents that may be audited.

3 Searchlight Architecture

3.1 Processing Model

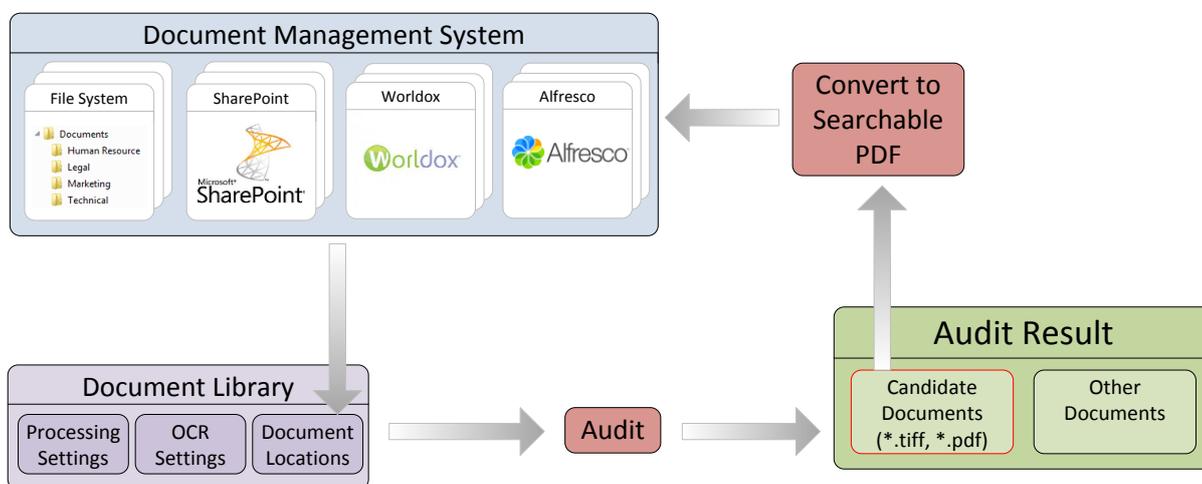


FIGURE 3.1: PROCESSING MODEL

3.1.1 Audit and Candidate Identification

Before processing a document library, Aquaforest Searchlight will perform an Audit (analysis) on the document library in order to determine which documents are candidates for processing by examining each document's searchability status (see next section) and the document library's processing settings.

3.1.2 Searchability Status

This determines whether a PDF file needs to be processed according to the Conversion Rules. Status values include Unknown, Searchable, Error, Partially Searchable amongst others.

3.1.3 Supported Formats

Aquaforest Searchlight currently supports only TIFF and PDF documents as input. As a result, candidate documents will always be of TIFF and/or PDF format.

3.2 Aquaforest Searchlight Windows Service

The service controls the execution of all job runs in Aquaforest Searchlight. It is used by the scheduler and enables the monitoring and processing of document libraries at regular time intervals without interfering with other work being performed on the machine it is installed in.

4 Document Library Management

Aquaforest Searchlight revolves around the concepts of document libraries, a document library is a set of documents which Aquaforest Searchlight monitors and ensures are converted where required in accordance with the Document Library Conversion Rules. This section will discuss the different concepts associated with Document Libraries.

4.1 Creating a Document Library

Creating Document Libraries in Aquaforest Searchlight is managed by a wizard. This wizard can be launched by clicking the "Create New Library" button on the Dashboard tab. The wizard contains five pages that will guide you through the set up.

The wizard provides helpful information throughout the different stages of the document library creation process which aids in better understanding the various steps and settings involved. Refer to sections on Library Settings and OCR Settings for detailed description of each of the settings.

4.1.1 Select Library Locations

Service Status : Running

This section enables users to specify the locations or URL of the documents they wish to monitor. To add a collection of documents to the document library follow the steps below.

- Select the Document Management system type you wish to add, this will be SharePoint or the Windows File System.
- When that is selected, a group box will be displayed that will allow the user to provide the location of the document collection and the security credentials (Username and Password).

- This will then display all the document collections on folders found at that particular location allowing the user to select which documents to add to the document library.
- If the documents are selected, the user should click on the “Add Another Location” button to add more documents to the library.
- If the User is happy with the documents added, a next button is provided to move to the next stage of the library set up.

4.1.2 SharePoint URL format

Below is an example of how to set the SharePoint URL format when setting up a document library in Searchlight.

SharePoint 2010, actual URL:

[http\(s\)://SharePoint2010/site/myLibrary/myForms/AllItems.aspx](http(s)://SharePoint2010/site/myLibrary/myForms/AllItems.aspx)

Supported URL format in Searchlight:

[http\(s\)://SharePoint2010/site/myLibrary](http(s)://SharePoint2010/site/myLibrary)

SharePoint 2013, actual URL:

[http\(s\)://SharePoint2013/site/Library/_layouts/15/start.aspx#/mylibrary/Forms/AllItems.aspx](http(s)://SharePoint2013/site/Library/_layouts/15/start.aspx#/mylibrary/Forms/AllItems.aspx)

Supported URL format in Searchlight:

[http\(s\)://SharePoint2013/site/Library/mylibrary](http(s)://SharePoint2013/site/Library/mylibrary)

4.1.3 Document Settings

This tab allows the user to specify more rules and criteria for the selection of documents to be added to the Document Library. These criteria include date created, date modified, file types to add, and document count limit.

4.1.4 OCR Settings

These tabs will allow the user set the OCR Engine to their taste. Aquaforest Searchlight comes bundled with two OCR Engines and each of these Engines has different settings.

4.1.4.1 Aquaforest OCR Engine Settings

This page allows the User to choose the settings suitable for the Aquaforest OCR engine.

4. OCR Settings

OCR Engine:
 Aquaforest Extended (IRIS)

General Settings

Auto Rotate
Off

Despeckle

Deskew
Off

OCR Language

Line Removal
Off

Box Graphics

MRC Compression
Off

Stamps

JBIG2 Compression
Off

Advance Flags

PDF/A1-b
Off

Continue On Error
Yes

PDF Specific Settings

DPI

Retain Metadata
No

Retain Bookmark
No

4.1.4.2 Extended OCR Engine Settings.

This page allows the User to choose the settings suitable for the extended OCR engine.

4. OCR Settings

OCR Engine:
 Aquaforest Extended (IRIS)

General OCR Settings

Auto Rotate Off	<input type="range"/>	Despeckle No Despeckle	<input type="text" value="0"/>	Select Language(s) <input checked="" type="checkbox"/> English <input type="checkbox"/> German <input type="checkbox"/> French <input type="checkbox"/> Spanish <input type="checkbox"/> Italian <input type="checkbox"/> British <input type="checkbox"/> Swedish <input type="checkbox"/> Danish <input type="checkbox"/> Norwegian <input type="checkbox"/> Dutch <input type="checkbox"/> Portuguese English
Deskew No	<input type="range"/>	DPI Auto	Interpolation Mode Normal	
Remove Dark Borders No	<input type="range"/>	JPEG Quality <input type="text"/>	Interpolation Value <input type="text"/>	
No OCR Off	<input type="range"/>	Work Depth <input type="text"/>	Advanced Despeckle No Despeckle	
Keep Original Image No	<input type="range"/>	PDF Version 1.4	Remove White Pixels No	
Continue On Error Yes	<input type="range"/>	Advanced Flags <input type="text"/>		

Advanced Pre-processing and Compression Settings

4.1.5 Scheduler

The scheduler allows Aquaforest Searchlight to automate the monitoring and running of document libraries. There are three types of schedule types namely:

- **Manual**
If this option is selected, Aquaforest Searchlight will start work on the document library whenever the user clicks on the start button. Note that Aquaforest Searchlight will process this document library only once per click.
- **Once per day**
This setting allows the user to select any time during the day that Aquaforest Searchlight will start processing the document library.
- **Continuous**
This allows a document library to be scheduled to run periodically between a start time and end time each day. The periods may be seconds, minutes or hours. For example, a document library may be specified to run every hour between 9:00 and 17:00.
- **Run Once**
This setting allows the user to schedule the job to run once on a given date and time.

5. Scheduler Settings

Manual

Once per day

At:
16 : 06

Continuous

Every:
1 Hours

Between 16 : 06 And 17 : 06

Run once

On:
12/06/2013 15

At:
16 : 06

4.2 Editing a Document Library

Aquaforest Searchlight allows the user to edit the library locations, library settings and the OCR Settings at any time. These settings are accessible through the Library Tab.

4.3 Deleting a Document Library

Aquaforest Searchlight will allow a user to delete a document library at any time as long as it is not in a running state.

4.4 Processing/Running a Document Library

Aquaforest Searchlight processes a document library by first carrying out an audit to find out how many files need to be converted. After a successful audit, Aquaforest Searchlight can then start conversion. Starting a run can either be done by the user or the scheduler. If the Ad-Hoc option is selected during the scheduler set up, the user will have to manually click on the "Run" button on the dashboard.

4.5 Review Audits & Conversions

Since Aquaforest Searchlight processes document in an in place fashion, it is best to always know the exact state of the Document Library before conversion begins. This process also helps the user compare the state of the Document Library before and after conversion.

The audit details are made available to user via the Run Details tab under the Library Tab. The audit details provide information such as: the list of all the files in the library, the number of searchable files in the library, the library size and much more.

A user can also choose the number of past Audits that should still be kept in Aquaforest Searchlight, this option is provided in the Library Settings tab both in the Create New Library wizard and under the Library Drill-Down tab.

4.6 Reports

Aquaforest Searchlight provides reports that show statistics for particular document libraries, conversions as well as error and support reports. Reports can be generated in PDF or Excel formats. The following reports are included with Aquaforest Searchlight.

- Aquaforest Searchlight Summary Report – Shows a summary of all the document libraries and settings (email settings, license, etc.) currently configured as well as the support status
- Document Library Report – Shows all the statistics related to a particular document library including the number conversions performed, success rate, documents processed, etc.
- Audit Report – Shows the results of an audit
- Conversion Report – Shows the statistics of a particular conversion
- Conversion Error Report – Shows all the errors and exceptions that occurred during a particular conversion

4.6.1 Error documents

With Aquaforest Searchlight, error documents are easy to locate. The Run Details tab provides filtering options which enables users to only display files using set criteria.

4.6.2 Library Status

This report can be accessed via the Library Status tab.

The screenshot shows the Aquaforest Searchlight interface. At the top, there is a navigation bar with 'Dashboard', 'Library', 'Settings', and 'Help & Support'. Below this, there are tabs for 'Status', 'Library Settings', 'Document Settings', 'OCR Settings', 'Run Details', 'Scheduler', and 'Alerts'. The main content area is split into two columns: 'STATISTICS' and 'LOG OUTPUT'. The 'STATISTICS' column shows data for PDF Documents, TIFF Documents, and Library Totals. The 'LOG OUTPUT' column shows the audit process details, including directory checks and document enumeration.

STATISTICS	
PDF Documents	
Total PDF Documents:	144
Image-only PDFs:	3 (2.1 %)
Partially Searchable PDFs:	5 (3.5 %)
Fully Searchable PDFs:	134 (94.4 %)
Error PDF Documents:	2
Total PDF Pages: 883	
Image-only Pages:	19 (2.2 %)
Fully Searchable Pages:	864 (97.8 %)
TIFF Documents	
Total TIFF Documents:	0
Error TIFF Documents:	0
Total TIFF Pages:	0
Library Totals	
Total Documents:	144
Total Error Documents:	2
Total Pages:	883
Total Searchable Pages:	864 (97.8 %)

```
LOG OUTPUT
Document Library ID: 1| Run ID: 464
02-Aug-2013 15:00:44: Starting Audit...
Enumerating documents...

Checking directory C:\Users\khalil\Desktop\Test Location\1\Samples_GDSI
Checking directory C:\Users\khalil\Desktop\Test Location\1
Checking directory C:\Users\khalil\Desktop\Test Location\1\Samples_Kolak
Checking directory C:\Users\khalil\Desktop\Test Location\1\Samples_Tab
Checking directory C:\Users\khalil\Desktop\Test Location\1\Set1
Checking directory C:\Users\khalil\Desktop\Test Location\1\Open Password
Checking directory C:\Users\khalil\Desktop\Test Location\1\PDF_Image_Types
Checking directory C:\Users\khalil\Desktop\Test Location\1\Set2
Documents enumerated (matching selection rules): 144

No new documents found to audit.
```

4.6.3 Generating Log files

Log files can be generated using the “Generate Log” button which can be found on the Run Details tab under the Library Tab. This enables users to capture the information generated by any given Run ID into a PDF file which can be sent to the support team. This provides a file by file assessment of a document library before and after a run is executed. It shows the files that have been converted successfully and those that have failed.

This following page illustrates the use of combo boxes to filter the results being displayed on the User Interface.

The screenshot shows the Aquaforest Searchlight interface. At the top, there are navigation tabs: Dashboard, Library, Settings, Help & Support. Below these are sub-tabs: Status, Library Settings, Document Settings, OCR Settings, Run Details, Scheduler, Alerts. The main content area displays a table with columns for RUN ID, RUN DATE, PROCESSING MODE, AUDIT RESULTS (Status, Successful Documents, Error Documents), and CONVERSION RESULTS (Status, Successful Documents, Error Documents). A 'Filter Status' dialog box is open, showing options: Completed (checked), Aborted, Not Yet Converted, and Service Error. Below the table, there are radio buttons for 'Audit' and 'Conversion', and a 'Total number of ru...' label. A detailed table shows document paths, audit status, file type, last modified, pages, and conversion status. At the bottom, there are navigation controls (Previous, Next, Limit 500) and a 'Total number of documents = 23' indicator.

RUN ID	RUN DATE	PROCESSING MODE	Status	Successful Documents	Error Documents	Status	Successful Documents	Error Documents
5	30/07/2013 12:43:	Audit and OCR	Completed	198	21	Complete	0	0
6	30/07/2013 12:49:	Audit and OCR	Completed	198	24	Complete	0	0
7	30/07/2013 12:53:	Audit and OCR	Completed	198	25	Complete	0	0

DOCUMENT PATH	AUDIT STATUS	FILE TYPE	LAST MODIFIED	PAGES	Deleted
C:\Users\khalii\Desktop\Test Location	searchable	.PDF	7/30/2013 12:40:35 PM	1	False
C:\Users\khalii\Desktop\Test Location	searchable	.PDF	7/30/2013 12:40:40 PM	1	False
C:\Users\khalii\Desktop\Test Location	searchable	.PDF	7/30/2013 12:40:46 PM	1	False
C:\Users\khalii\Desktop\Test Location	searchable	.PDF	7/30/2013 12:40:57 PM	1	False
C:\Users\khalii\Desktop\Test Location	searchable	.PDF	7/30/2013 12:40:46 PM	1	False
C:\Users\khalii\Desktop\Test Location	searchable	.PDF	7/30/2013 12:40:51 PM	1	False
C:\Users\khalii\Desktop\Test Location	searchable	.PDF	7/30/2013 12:40:56 PM	1	False
C:\Users\khalii\Desktop\Test Location	searchable	.PDF	7/30/2013 12:41:03 PM	1	False
C:\Users\khalii\Desktop\Test Location	searchable	.PDF	7/30/2013 12:41:03 PM	1	False
C:\Users\khalii\Desktop\Test Location	searchable	.PDF	7/30/2013 12:41:09 PM	1	False
C:\Users\khalii\Desktop\Test Location	searchable	.PDF	7/30/2013 12:41:10 PM	1	False

5 Microsoft SharePoint Document Stores

Aquaforest Searchlight can be configured to monitor multiple SharePoint libraries. Below are some concepts that should be taken into consideration during configuration.

5.1 Versioning

Since Aquaforest Searchlight uses in-place processing, the source document is replaced by the resulting PDF file. However, if versioning is turned on, the resulting PDF file will be created as another version of the input file in SharePoint. If versioning is turned off then the resulting PDF file replaces the source file and archiving is automatically activated.

5.2 Archiving

To avoid making inadvertent changes to the source document, Aquaforest Searchlight will ensure the archiving of source documents if versioning is turned off in SharePoint. Archiving is the process of copying over the source documents to an archive location specified by the user before performing any sort of processing on them.

5.3 Metadata

Metadata will be used to store the "Searchability" status of documents. Checks are made to determine if a document has been modified since it was last OCR'd and therefore may require re-processing.

5.4 Security

The user will have to provide the security credentials of the SharePoint library they wish to monitor during configuration.

6 File System Document Stores

PDF Search light can be configured to monitor folders on the windows file system. Below are a few issues that need to be considered when using the Windows File System.

6.1 File Name Length

The windows operating system has a limit to file name length it can process. Aquaforest Searchlight always runs an audit before any conversion is carried out. Before the start of an audit, if any files with long names are found it will be reported to the user and the user can either shorten the file names or move the files.

6.2 File Access Permissions

If there are any secured locations that are to be monitored, you will have to configure the Aquaforest Searchlight Service with the security credentials of a user that has permissions to access that particular location.

6.3 Archiving

To avoid any loss of document, Aquaforest Searchlight is created to enforce archiving of input document. This is the process of copying over input documents to an archive location specified by the user before processing them.

6.4 Metadata

Metadata will be held in the Aquaforest Searchlight database to store details of the searchability status of a document. Checks are made to determine if a document has been modified since it was last OCR'd and therefore may require re-processing.

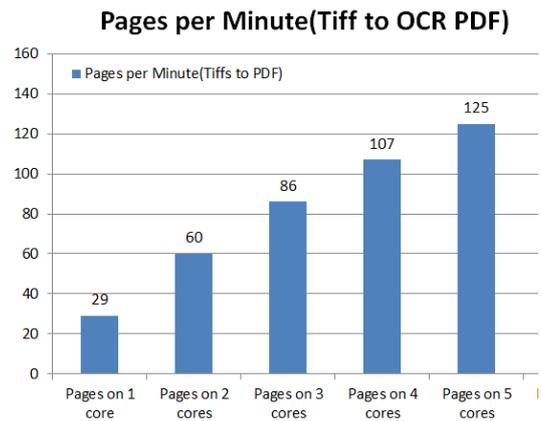
7 Aquaforest Searchlight Modules

7.1 Multi-Core Module (Included with the standard product)

This module is used to take full advantage of the number processors available on a computer.

The current release allows users to process up to 10 files in parallel thereby reducing the time needed to execute a job.

The chart gives some indication of the improvement in throughput that can be expected when using the multi-core module.



7.2 Extended (IRIS) OCR Module (Included with the standard product)

OCR engines are the components that perform the task of text recognition on image files and extraction. Aquaforest Searchlight ships two OCR Engines namely the Aquaforest OCR Engine and the Extended (IRIS) OCR Engine. Below is an explanation of the OCR Engines.

The Aquaforest OCR Engine is also included as a standard part of the product and can be used to convert Image PDFs and TIFFs to searchable PDF documents. This engine has support of about 24 European Languages, but you can only OCR using one language at a time.

The Extended Engine has the following benefits over and above the standard Aquaforest OCR engine:

- IRIS OCR Engine providing enhanced recognition
- New PDF Rasterizer component
- DOCX Output option
- Improved RTF Output
- CSV and Spreadsheet ML output options
- Supports over 100 Languages.
- Optional Asian Language Support
- Support for multiple languages within a single document from the same alphabet - e.g. French+ German + Italian

See section 8.3.5 for a detailed description of all the available Extended OCR settings.

7.3 Extended OCR Advanced Compression (Extra Cost Option)

Aquaforest Searchlight uses IRIS's New Intelligent High-Quality Compression (IHQC). IHQC offers the most impressive PDF colour compression without compromising visual quality, text resolution and legibility of your documents. The IHQC module will be available if you purchase the IHQC license. See section 8.3.6 for a detailed explanation of the settings available for Advanced Compression.

7.4 Metadata Extraction Module (Extra Cost Option)

Pingar Entity Extractor for SharePoint automatically enters keywords from the content of structured documents. With Pingar Metadata Extraction for SharePoint, you can automatically identify the following business entities within documents:

- Relevant Keywords
- Taxonomy Terms
- People
- Organizations
- Locations
- Addresses
- Phone Numbers
- Email Addresses
- URLs
- Dates
- Times
- Ages
- Money Amounts
- Credit Card Numbers
- Bank Account Information

For more information please contact support@aquaforest.com

7.5 Advanced iFilter Module (Extra Cost Option)

The freely available Adobe iFilter indexes PDF standard text well, but a more advanced solution is required for additional items such as metadata including titles, subjects, authors, keywords, PDF portfolios, annotations, text and data from forms, bookmarks, attachments, create time/date, number of page. Using such an iFilter can result in more precise search results.

For more information please contact support@aquaforest.com

8 Aquaforest Searchlight User Interface

8.1 Getting Started

When Aquaforest Searchlight is launched for the very first time, a getting started page is displayed to introduce the user to the different features of Aquaforest Searchlight and provide assistance in creating the first document library.

Welcome to Aquaforest Searchlight Version 1.04



Aquaforest Searchlight is able to monitor your SharePoint or File System document stores to ensure that all files are fully searchable.

To get started you will need to define a **Searchlight Document Library** that references the document store that you wish to monitor. You can process the Document Library in **Audit Mode** which will scan your documents and provide a report showing the number of files that are not fully searchable.

You can then process the library in **Make Searchable Mode** which will make use of OCR where required to make your documents fully searchable.

There is a sample Searchlight Document Library which you can process to get an understanding of how the product works and the [Reference Guide](#) provides more detailed information.



© Aquaforest Limited 2001-2014

Show this message on startup

Yes

[Continue](#)

8.2 Aquaforest Searchlight Dashboard

NAME	LIBRARY TYPE	LAST RUN	SCHEDULE	SEARCHABILITY	RUN STATUS
test one	File System	02-Aug-2013 17:28:11	Continuous	97.8 %	Completed

The dashboard gives a summary of the status of any document libraries currently processing or scheduled to process. A document library's previous conversions can be reviewed by selecting it from the list.

8.2.1 Dashboard Fields

Fields	Description
Name	Name of the document library
Library Type	The type of the document library: <ul style="list-style-type: none">• SharePoint• File System
Last Run	Time and date of the last run
Schedule	Manual or Automatic
% Searchable	The percentage of documents that is currently searchable in the document library
Status	Current status of the document library eg: <ul style="list-style-type: none">• Running• Completed• Error• Aborted
(Icons)	Abort, Pause, Start

8.3 Library

8.3.1 Library Status

This screen provides a detailed breakdown of all the document libraries currently configured in Aquaforest Searchlight. Each document library will have detailed information about each of the documents it contains and details about each document.

The screenshot shows the Aquaforest Searchlight interface. At the top, there is a navigation bar with 'Dashboard', 'Library', 'Settings', and 'Help & Support'. Below this, there are tabs for 'Status', 'Library Settings', 'Document Settings', 'OCR Settings', 'Run Details', 'Scheduler', and 'Alerts'. The 'Status' tab is active. The main content area is divided into two panels: 'STATISTICS' and 'LOG OUTPUT'. The 'STATISTICS' panel shows data for PDF and TIFF documents. The 'LOG OUTPUT' panel shows a list of directories being checked and the total number of documents enumerated.

STATISTICS	
PDF Documents	
Total:	291
Image-only:	2.2 %
Partially Searchable:	13 %
Fully Searchable:	84.8 %
Error:	22
Pages:	42849
TIFF Documents	
Total:	0
Error:	0
Pages:	0

```
LOG OUTPUT
Document Library ID: 1| Run ID: 1
13-Jun-2013 16:08:56: Starting Audit...
Enumerating documents...

Checking directory C:\qa\slqa\Bad
Checking directory C:\qa\slqa\Open Password
Checking directory C:\qa\slqa\PDF_Image_Types
Checking directory C:\qa\slqa\Samples_GDSI
Checking directory C:\qa\slqa\Samples_Kolak
Checking directory C:\qa\slqa\Samples_Tab
Checking directory C:\qa\slqa\Set1
Checking directory C:\qa\slqa\Set2
Documents enumerated (matching selection rules): 291
```

8.3.2 Library Settings

AQUAForest SEARCHLIGHT
Site3_Library3

Dashboard Library Settings Help & Support

Status Library Settings Document Settings OCR Settings Run Details Scheduler Alerts

Please enter a Library Name:

Library Type:

Locations: + Add new Location

↑
https://shrevin-hp/sites/sietcollectiontest/teamsite3/site3_library3

User Name	Password
Shrevin	●●●●●●●●

Choose Library Icon:

Processing Mode:
 Audit Only
 Audit and OCR

Cores:

SharePoint Settings

If Versioning is Off:

Publish Major Version:
 Yes

Check-In Comment:

Exclude Locations

🗑️ Delete
🔄 Refresh
💾 Save

Service Status : Stopped

Options	Description
Document Library Name	Name/Title/Description of the document library
Document Library Type	The type of the document library: <ul style="list-style-type: none"> SharePoint File System
Locations	One or more document locations to be processed.
Choose Library Icon	This is a button that allows users to check the file system for a library icon.
Processing Mode	<ul style="list-style-type: none"> Audit Only – analyse the document library to find out the documents that need to be converted without actually converting them Audit & OCR – perform audit on the document library and do convert the documents that have been identified as candidates for processing
Cores	This determines the maximum number of CPU cores that will be used when running the job.
SharePoint Versioning	This setting can be used to automatically turn versioning on.
Publish Major Version	Publish major version after OCR
SharePoint Exclude Locations	A set of URLs that will not be included in processing.

Check-in Comment	The check-in comment applied to the updated SharePoint file version.
------------------	--

8.3.3 Document Settings

PDF Document Selection

- Process PDF Documents: Yes
- Image Only PDFs: Yes
- Fully Searchable: Yes
- Partially Searchable: Yes
- Hidden Text: Yes

TIFF Document Selection

- Process TIFF Documents: Yes
- Delete Original TIFF: Yes

Document Archive Settings

- Archive Rule: Copy to Archive Folder
- Archive Template: %FILENAME%%TIMESTAMP%.%EXT%
- Archive Location: C:\TestFiles\Archive
- Archive source TIFFs to Archive Folder: Yes
- Archive source PDFs to Archive Folder (even when versioning is on): Yes

Document Error Settings

- Document Error Rule: Take no Action
- Document Error Location:

Filter Settings

- Filter Rule: No Filter
- From: 11/06/2014
- To: 11/06/2014

Advanced Settings

- Retry: No
- OCR Document Limit: 0
- Retain Creation Date: Yes
- Retain Modified Date: Yes
- Retain Created By: Yes
- Retain Modified By: Yes

Service Status: Stopped

Options	Description												
Document Selection Rule	The criteria to use to identify candidate documents for processing when auditing												
	<table border="1"> <thead> <tr> <th>Criteria</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>Process PDF</td> <td>Whether or not to process PDF documents</td> </tr> <tr> <td>Process TIFF</td> <td>Whether or not to process TIFF files</td> </tr> <tr> <td>Delete Original TIFF</td> <td>Whether or not to delete the original TIFF files after they have been converted to searchable PDFs</td> </tr> <tr> <td>Image Only</td> <td>Whether or not to process Image-only PDFs. An Image-only PDF is a PDF that originated from a scanned document or other digital image. An Image-only PDF does not contain any text, just pictures.</td> </tr> <tr> <td>Hidden Text</td> <td>Whether or not process PDF documents with hidden text in them</td> </tr> </tbody> </table>	Criteria	Description	Process PDF	Whether or not to process PDF documents	Process TIFF	Whether or not to process TIFF files	Delete Original TIFF	Whether or not to delete the original TIFF files after they have been converted to searchable PDFs	Image Only	Whether or not to process Image-only PDFs. An Image-only PDF is a PDF that originated from a scanned document or other digital image. An Image-only PDF does not contain any text, just pictures.	Hidden Text	Whether or not process PDF documents with hidden text in them
Criteria	Description												
Process PDF	Whether or not to process PDF documents												
Process TIFF	Whether or not to process TIFF files												
Delete Original TIFF	Whether or not to delete the original TIFF files after they have been converted to searchable PDFs												
Image Only	Whether or not to process Image-only PDFs. An Image-only PDF is a PDF that originated from a scanned document or other digital image. An Image-only PDF does not contain any text, just pictures.												
Hidden Text	Whether or not process PDF documents with hidden text in them												

		A Hidden Text PDF has pages that are Image-only with hidden (type 3) text. Such files are typically the output of running an OCR PDF process on an Image Only PDF.
	Partially Searchable	Whether or not to process PDF documents that are partially searchable, i.e., some pages are searchable and some are image-only
	Fully Searchable	Whether or not to process PDF documents that are fully searchable
	Date Created	Process documents that have been created between a defined time period
	Date Modified	Process documents that have been modified between a defined time period
	Retry	Whether or not to re-process documents that have previously failed to convert
	Limit	Limit the number of documents to process per run
Archive Folder	The folder location where original documents will be archived	
Archive Template	The template to use to rename the archived file name. The default is: %FILENAME%%TIMESTAMP%.%EXT%	
Archive source Tiffs to Archive folder	If enabled, this will Archive your source Tiffs to the Archive folder specified above.	
Archive source PDFs to Archive folder	If enabled, this will Archive the source PDFs to the Archive folder (even when versioning is enabled within SharePoint).	
Temp Folder	The location to use as the working folder. The default is: %TEMP%/pdfsearchlight	
Output File Name	The template to use to rename the original file name: The default is: %FILENAME.pdf	
Continue on Error	Whether or not to continue if an error occurs with a file in the document library	
Retain Creation Date	Retain the creation date of the source document (SharePoint creation date, FileSystem creation date and created date in PDF properties)	
Retain Modified Date	Retain the modified date of the source document (SharePoint modified date, FileSystem modified date and modified date in PDF properties)	
Retain Created By	Retain the created user of the source document (SharePoint created by, FileSystem owner and author in PDF properties)	
Retain Modified By	Retain the created user of the source document (SharePoint modified by)	

8.3.4 Aquaforest OCR Settings

Options	Description
Deskew	Straighten the image
Auto-Rotate	Automatically rotate pages so that text flows left to right
Despeckle	Remove specks below the specified pixel size from the image
Line Removal	This removes lines and boxes during OCR processing to improve recognition – particularly in cases where characters “touch” lines
Advanced Flags	Command line flags to be passed through to the underlying executable
Box/Graphics Processing	<p>By default, if an area of the document is identified as a graphic area then no OCR processing is run on that area. However, certain documents may include areas or boxes that are identified as “graphic” or “picture” areas but that actually do contain useful text.</p> <p>To ensure that the OCR engine can be forced to process such areas there are two options :</p> <p>“<i>Treat all Graphics Areas as Text</i>”. This option will ensure the entire document is processed as text.</p> <p>“<i>Remove Box Lines in OCR Processing</i>”. This option is ideal for forms</p>

	where sometimes boxes around text can cause an area to be identified as graphics. This option removes boxes from the temporary copy of the imaged used by the OCR engine. It does not remove boxes from the final image. Technically, this option removes connected elements with a minimum area (by default 100 pixels).
JBIG2 Compression	This option will compress bitonal images in generated PDFs using JBIG2 compression rather than the default Group 4 compression scheme. This will result in smaller PDF file sizes, at a cost of increasing processing time.
PDF/A1-b	Select Yes if the result file must be PDF/A1-b Compliant
MRC	This enables Mixed Raster Compression which can dramatically reduce the output size of PDFs comprising of colour scans.
DPI	Specify the DPI of the TIFF that will be used when converting TIFF to PDF. Set to Auto by default, alternatively can be set to 300, 200 or 150 to force a specific resolution.
Remove Hidden Text	Remove existing hidden text (text that was added as a result of a previous OCR) from the PDF file so that the resulting searchable PDF file does not have two layers of the same text.
Retain Metadata	Copy metadata from the source PDF to the Searchable result PDF
Retain Bookmarks	Copy bookmarks from the source PDF to the Searchable result PDF

8.3.5 Extended OCR Settings

AQUAFOREST SEARCHLIGHT

Dashboard [Library](#) [Settings](#) [Help & Support](#) Site3_Library3

[Status](#) [Library Settings](#) [Document Settings](#) [OCR Settings](#) [Run Details](#) [Scheduler](#) [Alerts](#)

OCR Engine:
 Aquaforest **Extended (IRIS)**

General OCR Settings

Auto Rotate Off	Despeckle No Despeckle	<input type="checkbox"/> Remove Blank Pages 0	Select Language(s) <input checked="" type="checkbox"/> English <input type="checkbox"/> Estonian <input type="checkbox"/> Faroese <input type="checkbox"/> Fijian <input type="checkbox"/> Finnish <input type="checkbox"/> French <input type="checkbox"/> Frisian <input type="checkbox"/> Friulian <input type="checkbox"/> Galician <input type="checkbox"/> Ganda <input type="checkbox"/> German
Deskew No	DPI Auto	Interpolation Mode Normal	
Remove Dark Borders No	JPEG Quality	Interpolation Value	
Keep Original Image No	Work Depth	Advanced Despeckle No Despeckle	
Remove Hidden Text No	PDF Version 1.4	Remove White Pixels No	
Remove Visible Text No	Advanced Flags		
Continue On Error Yes			

Advanced Pre-processing and Compression Settings

 Refresh
  Save

Service Status : Stopped

AQUAFOREST SEARCHLIGHT

Dashboard [Library](#) [Settings](#) [Help & Support](#) QA1

[Status](#) [Library Settings](#) [Document Settings](#) [OCR Settings](#) [Run Details](#) [Scheduler](#) [Alerts](#)

OCR Engine:
 Aquaforest **Extended (IRIS)**

Advanced Pre-processing and Compression Settings

Line Removal Settings		Binarization	iHQC Compression
Remove Lines No		Binarize No	iHQC Compression Off
Horizontal Clean X	Horizontal Max Thickness	Contrast	Quality Factor Medium
Vertical Clean X	Vertical Max Thickness	Brightness	Compression Level 1
Horizontal Clean Y	Horizontal Max Gap	Smoothing Level	
Vertical Clean Y	Vertical Max Gap	Threshold	
Horizontal Dilate	Horizontal Min Length		
Vertical Dilate	Vertical Min Length		

Options	Description
Autorotate	Detect page orientation and correct if required
Deskew	Rotates the image to correct its skew angle.
RemoveDarkBorders	Removes the dark surrounding from bitonal, grayscale or color images. The dark surrounding of the image is whitened. Note: The dark border should be touching the edge of the image/page for this to work.
Remove Hidden Text	Remove existing hidden text (text that was added as a result of a previous OCR) from the PDF file so that the resulting searchable PDF file does not have two layers of the same text.
Remove Visible Text	Avoid OCRing visible text (text as a result of conversion from an electronic document such as Word to PDF)
RemoveBlankPage	Set this to true to remove blank pages from Tiff or PDF documents. A value needs to be set for sensitivity (see below).
Sensitivity	The sensitivity, from 1 to 100. With a high sensitivity, less blank pages are detected.
Despeckle	Removes all the groups of connected pixels with a number of pixels below the parameter. Suggested range: 1-20.
Workdepth	This parameter (0 – 255) defines how deeply the OCR engine will analyze a page with 255 being the deepest. For poorer quality documents, higher values can give better recognition results.
JPEGQuality	This parameter (0 – 255) determines the compression/quality of color JPEG images in generated PDFs. 0 gives the smallest file size whilst 255 gives the best quality. The default value is 128.
PDFVersion	This determines the PDF version of the generated PDF : 1.4 1.5 1.6 1.7 PDF/A1B (PDF/A-1b)
Language	Determines the language to be used for OCR. English(Default) Faroese German Friulian French Greenlandic Spanish Haitian_Creole Italian Rhaeto_Roman British Sardinian Swedish Kurdish Danish Cebuano Norwegian Bemba Dutch Chamorro Portuguese Fijan

	Brazilian Galician Icelandic Greek Czech Hungarian Polish Romanian Slovak Croatian Serbian Slovenian Luxemb Finnish Turkish Russian Byelorussian Ukrainian Macedonian Bulgarian Estonian Lithuanian Afrikaans Albanian Catalan Irish_Gaelic Scottish_Gaelic Basque Breton Corsican Frisian Nynorsk Indonesian Malay Swahili Tagalog Japanese* Korean* Schinese* Tchinese* Quecha Aymara *Requires the Extended OCR Asian Module **No Language just a Latin alphabet	Ganda Hani Ido Interlingua Kicongo Kinyarwanda Malagasy Maori Mayan Minangkabau Nahuatl Nyanja Rundi Samoan Shona Somali Sotho Sundanese Tahitian Tonga Tswana Wolof Xhosa Zapotec Javanese Pidgin_Nigeria Occitan Manx Tok_Pisin Bislama Hiligaynon Kapampangan Balinese Bikol Ilocano Madurese Waray None** Serbian_Latin Latin Latvian
DPI	Sets the DPI of images in the output file. Set to Auto by default, alternatively can be set to 300, 200 or 150 to force a specific resolution.	
AdvancedDespeckle	The size of the speckles to remove.	
RemoveWhitePixels	By default, despeckle removes black pixels. If set to true, despeckle will remove white pixels rather than black pixels.	
Binarization	Whether or not to perform binarization on the document.	

Brightness	The brightness (higher values will darker the result).
Contrast	The contrast (lower values will darker the result).
SmoothingLevel	Smoothing may be useful to binarize text with a colored background in order to avoid noisy pixels (0 disables smoothing, higher values smooth more).
Threshold	Sets the threshold for fixed threshold binarization (0 for automatic threshold computation).
RemoveLines	Whether or not to remove lines from an image (The image must be black and white).
HorizontalCleanX	The parameter for cleaning noisy pixels attached to the horizontal lines.
HorizontalCleanY	The parameter for cleaning noisy pixels attached to the horizontal lines.
VerticalCleanX	The parameter for cleaning noisy pixels attached to the vertical lines.
VerticalCleanY	The parameter for cleaning noisy pixels attached to the vertical lines.
HorizontalDilate	The dilate parameter that helps the detection of horizontal lines.
VerticalDilate	The dilate parameter that helps the detection of vertical lines.
HorizontalMaxGap	The maximum horizontal line gap to close. It is useful to remove broken lines.
VerticalMaxGap	The maximum vertical line gap to close. It is useful to remove broken lines.
HorizontalMaxThickness	The maximum thickness of the horizontal lines to remove. It is useful to keep vertical lines larger than this parameter. Can be also useful to keep vertical letter strokes.
VerticalMaxThickness	The maximum thickness of the vertical lines to remove. It is useful to keep horizontal lines larger than this parameter. Can be also useful to keep horizontal letter strokes.
HorizontalMinLength	The minimum length of the horizontal lines to remove.
VerticalMinLength	The minimum length of the vertical lines to remove.
RemoveDarkBorders	Removes the dark surrounding from bitonal, grayscale or color images. The dark surrounding of the image is whitened (Note: The dark border should be touching the edge of the page for this to work).
Interpolation	Interpolates the source image to the given resolution. This value (the target resolution) must be greater than the

	source image's resolution.
InterpolationMode	Sets the interpolation mode.
KeepOriginalImage	Yes to keep the original image as it is. No to output the image generated after pre-processing is applied.

8.3.6 Advanced Compression Settings

The following parameters are needed to use the advanced compression settings (IHQC – Intelligent High Quality Compression)

Options	Description
MRC	Apply intelligent High Quality Compression. True or False
IHQCLevel	The compression level to be used. Level 1 is the basic compression level. Level 3 is the most advanced intelligent High Quality Compression mode. Compression Level 3 2b 2a 1
IHQCQF	The IHQC quality factor: Quality Maximal quality Very high quality High quality Favour quality over size Medium Favour size over quality Small size Very Small Size Minimal size

8.3.7 Scheduler Settings

Dashboard Library Settings Help & Support QA1

Status Library Settings Document Settings OCR Settings Run Details **Scheduler** Alerts

Manual

Once per day
At: 16 : 29

Continuous
Every: 1 Hours
Between 16 : 29 And 17 : 29

Run once
On: 13/06/2013 15
At: 16 : 29

Options	Description
Stop Processing on Error	If checked, the job will stop if it returns an error, and will not run again until the error is cleared.
Send Email on Error	If checked, Aquaforest Searchlight will send an email if an error occurs with a document library. The email server configuration should be setup – see the email configuration section for more information.
From Email Address	The “from” email address that will be used for the message.
To Email Address	The email address that the message will be sent to.
Email Title	The title of the email.
Email Message	The body of the email.
Ad-Hoc (By User)	This option implies the Aquaforest Searchlight will convert files only when the user clicks the run button
Run Once At	This option allows a user to specify the exact date and time that the job should be run
Once a day	This option allows the user to specify a particular time daily that the job should run.
Continuous	The continuous option allows the job to run on a fixed regular schedule.

8.3.8 Alert Settings

8.3.9 Run Details

RUN DATE	Status ▼	AUDIT		CONVERSION		
		Successful Documents	Error Documents	Status ▼	Successful Documents	Error Documents
05/06/2013 15:47:16	Stopped	272	23	Not Yet Converted	0	0
05/06/2013 16:07:54	Stopped	0	0	Stopped	21	2

DOCUMENT PATH	AUDIT STATUS ▼	FILE TYPE ▼	LAST MODIFIED ▼	PAGES	CONVERSION STATUS ▼
C:\qa\slqa\Bad\5844bbb2-9c57-11dd-83db-9d068457f227.pdf	error	.PDF	6/15/2009 11:09:35 AM	0	Not Yet Converted
C:\qa\slqa\Bad\aaasdad.pdf	error	.PDF	4/29/2009 9:59:46 AM	0	Not Yet Converted
C:\qa\slqa\Bad\ABAKHOUR .pdf	error	.PDF	10/11/2007 12:51:16 PM	0	Not Yet Converted
C:\qa\slqa\Bad\ABI-LOTF.pdf	error	.PDF	10/11/2007 12:51:16 PM	0	Not Yet Converted
C:\qa\slqa\Bad\birneywimage2.pdf	error	.PDF	10/11/2007 1:05:47 PM	0	Not Yet Converted
C:\qa\slqa\Bad\example.pdf	error	.PDF	10/11/2007 1:06:52 PM	0	Not Yet Converted
C:\qa\slqa\Bad\Fax00000000_1.pdf	error	.PDF	8/28/2009 5:08:30 PM	0	Not Yet Converted
C:\qa\slqa\Bad\Functioneel Ontwerp Sharepoint 2003.pdf	error	.PDF	4/29/2009 9:49:47 AM	0	Not Yet Converted
C:\qa\slqa\Bad\Legacy-Smith 00678.pdf	error	.PDF	8/10/2010 1:05:18 AM	0	Not Yet Converted
C:\qa\slqa\Bad\out_merge.pdf	error	.PDF	11/4/2008 5:31:51 PM	0	Not Yet Converted
C:\qa\slqa\Bad\p1000.pdf	error	.PDF	12/22/2008 2:15:24 PM	0	Not Yet Converted

The Run History list box represents a list of the previous runs carried out on the Document Library. Filters can be applied to limit the runs and individual files shown.

The Generate Log Report button is used to generate the log report of the selected run/documents as a pdf file.

The user can right click on any file in the run details data grid to open the file, open the file location and also view the log entry associated to the particular file.

8.3.10 Run Details Context Menu

Using the Right-Click context menu on a particular file allows the file or file location to be opened.

DOCUMENT PATH	AUDIT STATUS	FILE TYPE
C:\qa\slqa\Bad\5844bbb2-9c57-11dd-83db-9d068457f227.pdf	error	.PDF
C:\qa\slqa\Bad\asdad.	error	.PDF
C:\qa\slqa\Bad\ABAKH	error	.PDF

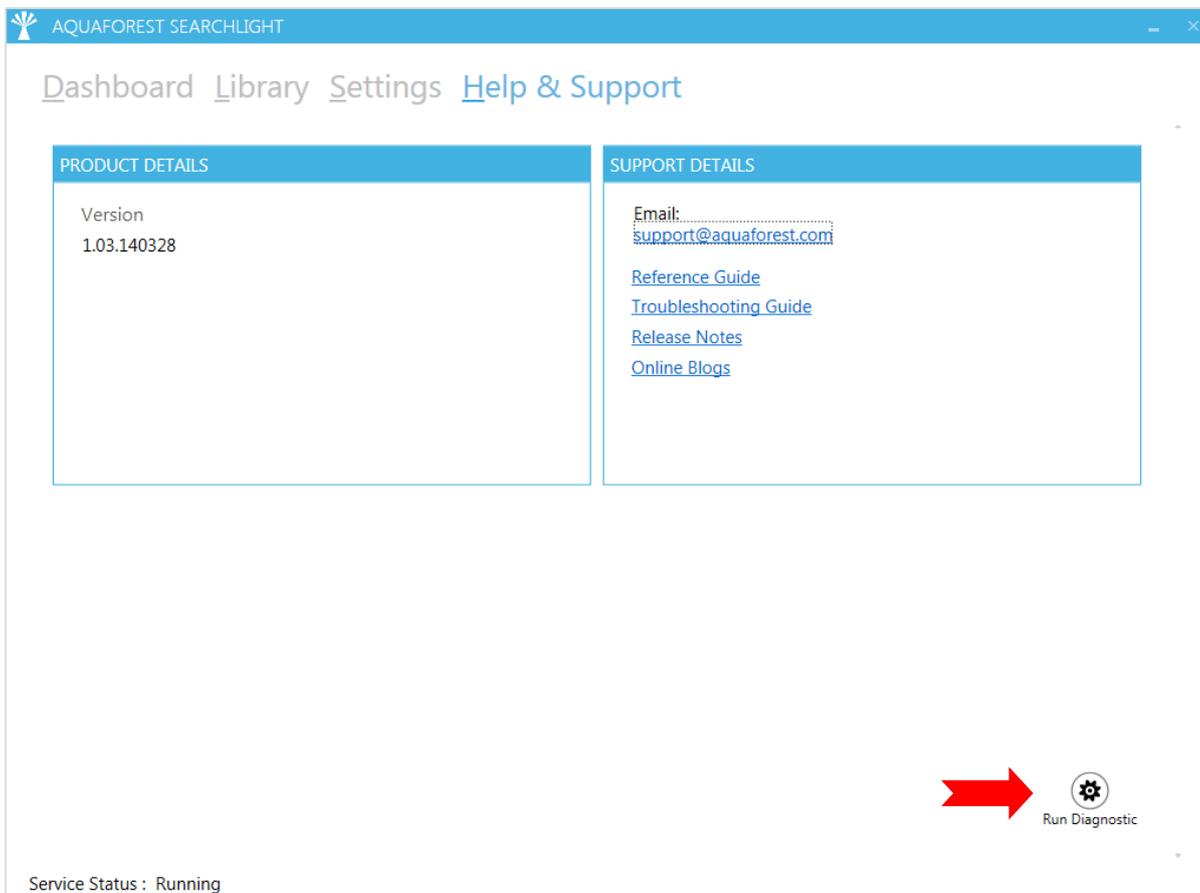
Context menu options: Open File, Open File Location

8.4 Support

The support page is the starting point for help with Aquaforest Searchlight. It provides resources such as the reference guide, release notes and online blogs. It also provides the generic support email address which should be used in the first instance when reporting an issue or any queries.

8.5 Diagnostic Tool

In order to run the diagnostic tool, click on the “Run Diagnostic” button in the “Help & Support” tab as pointed out in the image below. This will initiate the diagnostic wizard which will run various checks to determine if your system meets all the requirements needed to run Aquaforest Searchlight as well as collect information related to a particular document library. All the gathered information will be made available in a zip file which can be sent to support@aquaforest.com for further investigation.



8.6 Settings

A AQUAForest SEARCHLIGHT

[Dashboard](#)
[Library](#)
[Settings](#)
[Help & Support](#)

EMAIL SETTINGS

SMTP Server

SMTP Port

Username

Password

Re-enter Password

↻
F

Refresh Save

LICENSE

License Type: Permanent

Computer Bound: No

Computer Identifier:

Document Limit: Unlimited

Features: Aquaforest OCR; Extended OCR; Extended OCR Asian; Multi-Core

License Key:

F7FE5D9F1EA63C7153DBE25A68FF498C8F38241AD8EE48433C0
 3D424A762CA7D8D643A992B34997A7F2100A084D276D242570
 E9BB5E74906977B2527B4DD5D90CD719BA7E0B0880281F6ED0
 98EE62CD7F579905BF7CF52CA9A612B46E53E77D49323F478B1
 4B9A9FC8CA6058CCBD102875C2B3C6E23C4F5938E46E8586C8

F

Update

CHANGE THEME

Light Blue

Dark Blue

Light Red

Dark Red

Light Purple

Dark Purple

Light Green

Dark Green

Light Orange

Dark Orange

Show Welcome Screen on start-up

No

Service

On

Service Status : Running

8.6.1 License Settings

Options	Description
License Type	Is the license Temporary or Permanent
Computer Bound	Is the license Computer Bound or Non-Computer Bound
Document Limit	Document conversion limit for Temporary License key
Features	Modules enabled by the current license
License Key	License Key installed

8.6.2 Email Settings

The settings screen allows email server information to be defined. This is used to support the "Send Email on Error" functionality defined in the Scheduler Settings. The password is stored in an encrypted form.

Options	Description
Server	Address of the Server hosting the SMTP server.
SMTP Port	SMTP Server port. Default is 25.
Username	Username for authentication by the server.
Password	Password for the username.
Need Authentication	If checked, authentication will be used and the username and password can be filled in.

8.6.3 Themes

Colour Themes to choose from for the User Interface. The Light Blue theme is selected by default.

9 Aquaforest Searchlight Installation Files

The installation folder (by default C:\Aquaforest\Searchlight) contains the following :

Bin

This folder contains all the executables and libraries that searchlight uses manage the document libraries.

Config

This folder contains the configuration file. This should only be modified on guidance from the support team.

Data

Searchlight uses a sql lite database file named "searchlight.db" to store the details, conversions and audit history. We also provide a SQLite Client folder which contains a GUI that you can use to access and view the present state of the database. Note: Editing the database can cause unwanted behaviour by searchlight.

It is a good practice to make regular backups of the searchlight.db.

Docs

This folder contains reference guides and release notes for aquaforest searchlight.

Extended OCR

This folder contains the executables for the I.R.I.S Extended OCR engine.

License

This folder use to store the license key of searchlight and the license agreement information.

Live

The live folder contains three folders that are used to store the audit and conversion details at runtime.

Livelog

This folder is used to store the live OCR log information of the files that are currently being converted, the log files are grouped into folders by their document library id.

Log

The log folder is used to store the log details of the whole run from auditing to conversion. The log files are grouped into folders by their document library id.

Stats

This folder contains the document library statistics stored in xml files, each document library has an associated xml file with the same name as its id. These file contains a breakdown of searchable, Non searchable documents.

Temp

Before searchlight performs conversions on any pdf document, it copies the files to the temp/source directory and the output is stored in the tem/target directory. After a successful conversion, the file is copied back to its original location.

Searchlight also download and process sharepoint files (both tiff and pdf) to the temp folder, after the processing is complete, the files are uploaded.

Tj

This folder contains the executables for the aquaforest OCR engine.

10 Acknowledgements

This product makes use of a number of Open Source components which are included in binary form. The appropriate acknowledgements and copyright notices are given below.

LEPTONICA

Copyright (C) 2001 Leptonica. All rights reserved.

LIBJPEG

This software is based in part on the work of the Independent JPEG Group.

ZLIB

(C) 1995-2004 Jean-loup Gailly and Mark Adler.

ITEXT 4.1.6

Copyright (C) 1999-2009 by Bruno Lowagie and Paulo Soares et al. All Rights Reserved. Binaries distributed under the Mozilla Public License.

CUNEIFORM

Copyright (c) 1993-2008, Cognitive Technologies. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution. Neither the name of the Cognitive Technologies nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE

LIBTIFF

Copyright (c) 1988-1997 Sam Leffler. Copyright (c) 1991-1997 Silicon Graphics, Inc.

Permission to use, copy, modify, distribute, and sell this software and its documentation for any purpose is hereby granted without fee, provided that (i) the above copyright notices and this permission notice appear in all copies of the software and related documentation, and (ii) the names of Sam Leffler and Silicon Graphics may not be used in any advertising or publicity relating to the software without the specific, prior written permission of Sam Leffler and Silicon Graphics.

THE SOFTWARE IS PROVIDED "AS-IS" AND WITHOUT WARRANTY OF ANY KIND, EXPRESS, IMPLIED OR OTHERWISE, INCLUDING WITHOUT LIMITATION, ANY WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL SAM LEFFLER OR SILICON GRAPHICS BE LIABLE FOR ANY SPECIAL, INCIDENTAL, INDIRECT OR CONSEQUENTIAL DAMAGES OF ANY KIND, OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER OR NOT ADVISED OF THE POSSIBILITY OF DAMAGE, AND ON ANY THEORY OF LIABILITY, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

FREEIMAGE

This software uses the FreeImage open source image library. See <http://freeimage.sourceforge.net> for details. FreeImage is used under the (FIPL), version 1.0.