

Worldox with Autobahn DX

Ensure all Worldox Documents are Fully Text-Searchable

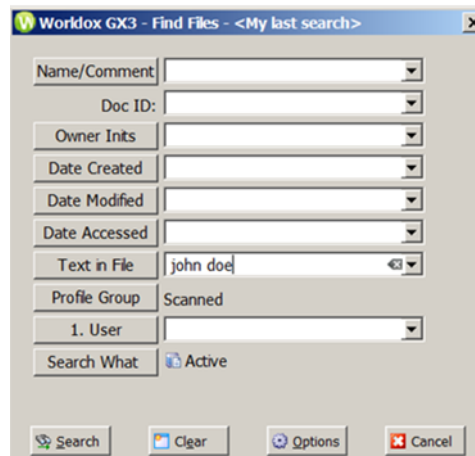


- ✓ High Performance OCR
- ✓ Convert Scanned PDF and TIFFs
- ✓ 24 Languages Supported
- ✓ Compressed Searchable PDFs
- ✓ Automated Job Processing
- ✓ No Page Count Limits

Overview

The Worldox document management system is widely used in the Legal industry but does not come with any in-built OCR capabilities. This means that any scanned PDF (“image-only PDF”) files or TIFF files will not be searchable via the Worldox text search tool.

Autobahn DX can quickly be configured to act as a “monitor” for Worldox to ensure that all image-only PDF files will be automatically converted to fully text-searchable PDF. In addition, TIFF files can be converted to fully searchable PDF



Processing will be applied to both newly added files and existing documents to ensure that text search is fully enabled.

The brief guide below is intended to help administrators configure Autobahn DX appropriately.

Autobahn DX

Autobahn DX is an OCR and document conversion processing product designed to fit into an organization's document workflow. It is able to process and convert a variety of different types of document including TIFF images, Microsoft Office documents and HTML pages, convert those documents to searchable PDFs and perform a variety of other customizable activities.

Aquaforest Limited
Kingfisher Exchange
Walton Street
Aylesbury, BUCKS HP22 4PX

Phone: +44 1296 468168
Web : www.aquaforest.com
E-mail: info@aquaforest.com



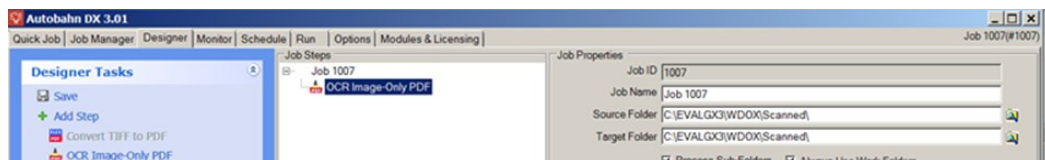
Configuring Autobahn DX for Worldox

Step 1

Determine the folder(s) to be monitored – for example C:\EVALGX3\WDOX\Scanned\

Step 2

Create a new Autobahn DX job that will monitor the Worldox folder to convert Image PDFs to fully text searchable PDFs.



Step 3

Configure the job using the notes below as a guide. The Autobahn DX reference guide contains a full description of the available settings.

1 Source and target folders should be the same.

2 Subfolders should be processed

3 For audit purposes the original files should be copied to an archive folder

4 To avoid re-processing files and to limit how many files can be processed in a single batch

5 This limits the files to be converted to unprocessed PDFs, with a limit of 1000 per batch*

6 The output file name is configured to overwrite the original non-searchable version

Job Properties	
Job ID	1007
Job Name	Job 1007
Source Folder	C:\EVALGX3\WDOX\Scanned
Target Folder	C:\EVALGX3\WDOX\Scanned
<input checked="" type="checkbox"/> Process Sub-Folders	<input checked="" type="checkbox"/> Always Use Work Folders
Input Files	Copy to Archive after Processing
Rename Input Files	%FILENAME%TIMESTAMP%.%EXT%
Filter Files	Include Unprocessed PDFs Only - with Document Count Limit
Filter Expression	.pdf,100
Files in Error	Copy to Error Folder <input type="checkbox"/> Retain Folder Structure
Log Files	
Log File	C:\Aquaforest\Autobahn DX 3.01\logs\1007%TIMESTAMP%.txt
CSV Log File	C:\Aquaforest\Autobahn DX 3.01\logs\1007%TIMESTAMP%.csv
Processing Folders	
Archive Folder	C:\Aquaforest\Autobahn DX 3.01\work\Archive\1007
Work Folder	C:\Aquaforest\Autobahn DX 3.01\work\1007
Error Folder	C:\Aquaforest\Autobahn DX 3.01\work\1007\errors
Temp Folder	C:\Aquaforest\Autobahn DX 3.01\work\1007\temp
Step Properties	
1 - Conversion Settings	
Output File Name	%FILENAME.pdf
Create Directories if Required	Yes
Continue on Error	Yes
OCR	Yes
OCR Language	English
Deskew	No
Auto-Rotate	No
Despeckle	0
OCR to TextFile	No
Output File	Plain Text
Extract PDF Images Via	Auto
Non-Image PDFs	OCR
Retain Bookmarks	No
Retain Metadata	No
Advanced Flags	
Output File Name	
Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of original file)	

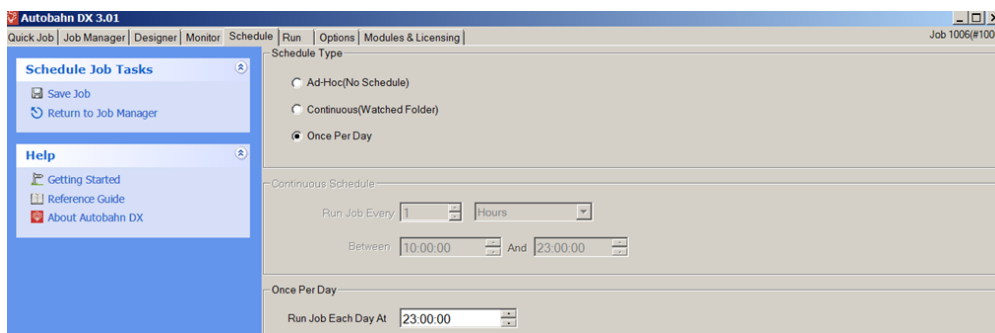
*This gives an upper limit on the number of documents that will be processed in a single batch run. Depending upon the document store and the schedule a larger or smaller figure may be appropriate.

Configuring Autobahn DX for Worldox (continued)

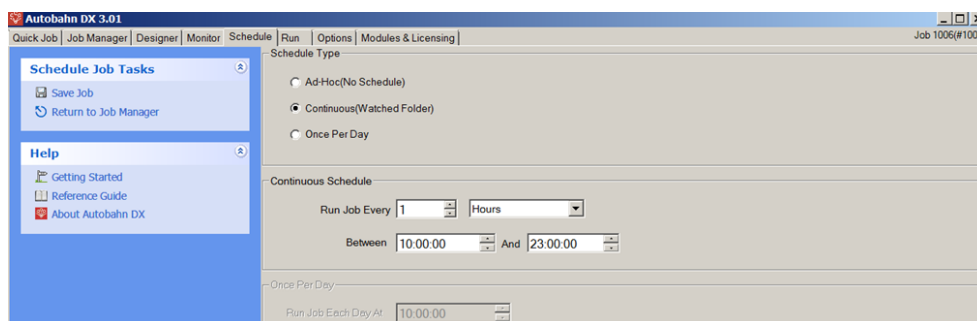
Step 4

Determine the job schedule and set accordingly. Jobs may be configured as Ad-Hoc which means that they will only run when manually initiated.

For automatic jobs there are two options – once per day which allows specification of a time to run. This is often configured to run outside of business hours in order to avoid impacting on system performance. For installations with a large number of documents this is generally the preferred approach.



The other option is to run the job in Continuous (Watched Folder) mode and determine how often the job should run.



Support and Further Information

Support is available from Aquaforest via email at support@aquaforest.com or by phone on +44 (0)1296 46816

More detailed product information is available at Aquaforest's web site :

Autobahn DX Product Page : <http://www.aquaforest.com/en/autobahn.asp>

Autobahn DX Reference Guide : <http://www.aquaforest.com/files/autobahndx30.pdf>

Autobahn DX Service Packages : http://www.aquaforest.com/en/autobahn_services.asp

Video Tutorial : http://www.aquaforest.com/video/adx_video_tutorial.htm