

# Aquaforest

Making SharePoint  
Content Findable

 Office 365

 SharePoint

 Microsoft Azure



## Aquaforest Searchlight

Automated Metadata and OCR



Tags



OCR PDF



Taxonomy



Entities



Rules



Zonal



Forms



Barcode



Dramatically Improve your Search with  
Aquaforest Searchlight.

# Before Automated Metadata

Office 365 | SharePoint | **Aquaforest**

BROWSE

ESPC17 Presentation EDIT LINKS

## Before Library

Search...

Home  
Notebook  
Documents  
Tasks  
Calendar  
Recent  
source files  
After Library  
Demo OCRd Library  
Demo Library  
**Before Library**  
Before Library  
Demo Image  
After Search  
Before Search  
Demo Search  
After Library  
Site contents  
EDIT LINKS

Type	Name
	Fillcancelation
	Final Form
	FIRA2015CFP0311
	icra06
	Patent Form
	Patent Recall
	Robotics Presentation
	Robotics
	robotics_automationtech
	Robots-Req1-3
	Telerobotics Concept Paper
	US20020157388
	US20090137952
	US20100258362
	US20110196798
	US7590680

+ Add document

The metadata columns have a lot of empty values.

Title	Patent ID	Keywords	People NLP
PowerPoint Presentation			
Artificial Intelligence & Robotics			
Robotics			
VTE Framework: Robotics & Automation Technology			
Robots In Manufacturing			
[Document Title]			

# After Automated Metadata

Office 365 | SharePoint

BROWSE | FILES | LIBRARY

ESPC17 Presentation EDIT LINKS

## After Library

Home

Notebook

Documents

Tasks

Calendar

Recent

- source files
- After Library
- Demo OCRd Library
- Demo Library
- Before Library

Before Library

Demo Image

After Search

Before Search

Demo Search

After Library

Site contents

EDIT LINKS

Type	Name	Title	Patent ID	Keywords	People NLP
PDF	Fillcancelation	Fill Cancellation	US 20020157388A1	World Intellectual Property Organization; Patent and Trademark Office; Intellectual property	Anna Friesel; Maura McMahon; Charles Lamarche; Chen-Chien James Hsu; Yong Hong; Christine Shaw
PDF	Final Form	Final Form	US20020157388	World Intellectual Property Organization; Patent and Trademark Office; Intellectual property	Maura McMahon; Charles Lamarche; Pat McColl; Christine Shaw; Bruce Kennen
PDF	FIRA2015CFP0311	Federation of International Robot-soccer Association		Collaborative learning; Education; Google Scholar; Architectures; Mathematica; Pattern matching; Technologies; Perception; Cognitive robotics	Jean Marc Thiriet; Yang Tu; Marian Poboroniuc; Jae Ryoo; Juan Carlos Rosete Fonseca; Joseph Fourier Grenoble; Anna Friesel; Yong Hong; Chen-Chien James Hsu; Dorian Cojocaru; John Anderson; Zbigniew Mrozek; Guido Bugman; Dante Barone; Chung-Hsien Kuo
PDF	icra06	PowerPoint Presentation		Constraints; Artificial intelligence; Approximation; Robustness; PowerPoint; Placement	Etienne Burdet; Brian J. Sauser; Wang Fei; Charles Lamarche; Rodney Brooks; Jin Sung Physical; David Kempeskie; Jean-Claude Latombe
PDF	Patent Form	Patent Form	US 20100258362A1	World Intellectual Property Organization; Patent and Trademark Office; Intellectual property	John Kennedy; Jae Ryoo; Marian Poboroniuc; Maura Banta; Jean Marc Thiriet; Mark D
PDF	Patent Recall	Patent Recall	US00759068082	World Intellectual Property Organization; Patent and Trademark Office; Intellectual property	Joseph P.Fernando; John McDonagh; Rodney Brooks
PDF	Robotics Presentation	Artificial Intelligence & Robotics		Constraints; Engineering; Artificial intelligence; Automation; Sensors and actuators; Intelligent agents	How; Mark D; John Kennedy; Maura Banta
PDF	Robotics	Robotics		Technologies; Experimentation; Performance; Computer vision; Architectures; Security services; Sensors and actuators; Supervised learning; Neural networks; Speech recognition; Reinforcement learning; Uncertainty; Programming interfaces; People with disabilities; Automation	Karel Capek
PDF	robotics_automationtech	VTE Framework: Robotics & Automation Technology		Technologies; Transportation; Performance; Organizations; Engineering; Licensing; E-learning; Marketing; International Organization for Standardization; Process control systems	Bruce Kennen; Pat McColl; Maura McMahon; Christine Shaw; Charles Lamarche; Marnie Jain; Cecilia Smith; Rebecca Buck; David Kempeskie; Vanessa Calderón-Rodríguez; John Kennedy Martin

Search this site

The metadata columns have been populated after running Searchlight through the library

# Before Searchable PDF Processing

The screenshot shows the SharePoint search interface. At the top, there is a purple header bar with the 'Office 365' and 'SharePoint' logos on the left, the 'Aquaforest' logo in the center, and a notification bell icon on the right. Below the header, on the left, is a blue sidebar with the SharePoint logo and the text 'We don't have any refiners to show you'. To the right of the sidebar is a blue button labeled 'Before Library'. The main search area has a search bar containing the text '"Patent Application Publication"'. Below the search bar are tabs for 'Everything', 'People', 'Conversations', and 'Videos'. A dropdown menu shows 'Preference for results in English'. The search results are displayed as two entries, each with a PDF icon and a title: 'US20100258362' and 'US20090137952'. Each entry includes a brief description and a link to the document. At the bottom of the search results, there is a blue box with the text 'Search Results for a term returns 2 results.' and an arrow pointing up to the search results area. To the right of the search results, there are links for 'Alert Me', 'Preferences', and 'Advanced Search'.

Office 365 | SharePoint | Aquaforest

Before Library

"Patent Application Publication"

Everything | People | Conversations | Videos

Preference for results in English ▼

**US20100258362**  
US 20100258362A1 (s) United States (12) **Patent Application Publication** (io) Pub ...  
Dgfe: Oct. 14s 2010 (54) ACTUATOR POWERED DEFORMABLE SOFT-BODIED  
AUTONOMOUS PLATFORMS (76) Inventor ...  
[aquaforest.sharepoint.com/ESPC17 Presentation/.../US20100258362.pdf](#)

**US20090137952**  
US 20090137952AI (s) United States (12) **Patent Application Publication** (io) Pub ... 2  
of f 64 US 2009/0137952 A1 **Patent Application Publication** ay Fig 2APatent  
Application Publication ...  
[aquaforest.sharepoint.com/ESPC17 Presentation/.../US20090137952.pdf](#)

2 results

Alert Me | Preferences | Advanced Search

Search Results for a term returns 2 results.

Aquaforest

# After Searchable PDF Processing

SharePoint **Aquaforest**

After Library

After running Searchlight the search results for the same term returns 3 results.

Patent ID

- US 20020157388A1 (1)
- US 20090137952A1 (1)
- US 20100258362A1 (1)

Keywords

- Batteries (1)
- Functionality (1)
- Interconnect (1)
- Measurement (1)
- Packaging (1)
- SHOW MORE


People


- Barry Trimmer (1)
- Bhaskar S. Ramamurthy (1)
- Neal A. Tanner (1)
- Randall L. Schlesinger (1)
- Takeshi Seto (1)


"Patent Application Publication"

Everything People Conversations Videos

Preference for results in English ▼

 **ACTUATOR POWERED DEFORMABLE SOFT-BODIED AUTONOMOUS PLATFORMS**  
US 20100258362A1 (\*) United States (12) **Patent Application Publication** (io) Pub ... Dgfe: Oct. 14s 2010 (54) ACTUATOR POWERED DEFORMABLE SOFT-BODIED AUTONOMOUS PLATFORMS (76) Inventor ...  
[aquaforest.sharepoint.com/ESPC17 Presentation/.../US20100258362.pdf](#)

 **PUMP-INTEGRATED FLEXIBLE ACTUATOR**  
Us 20020157388A1 (\*) United States (12) **Patent Application Publication** (10) Pllb ... BA 42 44 14 34 34 88 10 [**Patent Application Publication** Oct. 31, 2002 Sheet 5 of 9 US 2002I0157388 ...  
[aquaforest.sharepoint.com/ESPC17 Presentation/.../US20020157388.pdf](#)

 **ROBOTIC INSTRUMENT SYSTEMS AND METHODS UTILIZING OPTICAL...**  
US 20090137952A1 (\*) United States (12) **Patent Application Publication** (io) Pub ... 2 of f 64 US 2009/0137952 A1 **Patent Application Publication** ay Fig 2APatent Application Publication ...  
[aquaforest.sharepoint.com/ESPC17 Presentation/.../US20090137952.pdf](#)

3 results

Alert Me Preferences **Advanced Search**

**Aquaforest**

# How does it work?

# Aquaforest

Making SharePoint  
Content Findable

 Office 365

 SharePoint

 Microsoft Azure



## Aquaforest Searchlight

Automated Metadata and OCR



Tags



OCR PDF



Taxonomy



Entities



Rules



Zonal



Forms



Barcode





# Zonal Extraction



# Entity Extraction



Find on page Barry Trimmer

1 of 23

**Trimmer** (43) **Pub. Date:** **Oct.**

(54) **ACTUATOR POWERED DEFORMABLE SOFT-BODIED AUTONOMOUS PLATFORMS**

(76) Inventor: **Barry Trimmer**, Medford, MA (US)

(51) **Int. Cl.**  
**B25J 11/00** (2006.01)  
**G06F 19/00** (2006.01)

(52) **U.S. Cl.** ..... 180

pdf

Searchlight will extract text from input files.

**STATISTICS**

Statistics for Entire Job

Total Processes:	389
- Successful:	294
- Partially successful:	32
- Failed:	63
Total Metadata Tagged:	2436
- NLP:	637

**LOG OUTPUT**

Processing '<https://aquaforest.sharepoint.com/ESPC17%20Presentation/Test/US20100258362.pdf>'  
27-Oct-2017 14:54:47: Start Time

Tika text extraction: 1615 ms

[Text Comparison] - Successful

[NLP Extraction] - Successful  
- Barry Trimmer

[PDF Metadata Extraction] - Successful  
No PDF metadata found.

Aquaforest Searchlight

Searchlight will then extract entities from the text using NLP services such as: Open Calais, Rosette and Microsoft Cognitive Services.

Type	Name	Title	Patent ID	Keywords	People NLP
Patent Form	NEW		US 20100258362A1		
US20100258362A1	ACTUATOR POWERED DEFORMABLE SOFT-BODIED AUTONOMOUS PLATFORMS		US	Electromechanical system; Non-volatile memory; Electromechanical systems; Logic circuits; Flash memory; wireless device; Wireless devices; Logic circuit; transducer; Transducers; Interconnect; simulation; functionality; micro-electromechanical systems	Barry Trimmer

Searchlight will then add the extracted entities as values to metadata columns in SharePoint.

# Barcode Values



**ZONE DEFINER**

1 (of 19) 100%

**United States**

**Patent**

**Seto et al.**

**US 20020157388A1**

**US20020157388**

**PUMP-INTEGRATED FLEXIBLE ACTUATOR**

**US 20020157388A1**

**ABSTRACT**

A pair of flexible actuators (4) according to the present invention is supported by a fixture (2) made of a hard rubber at base ends of the flexible actuators (4). Each flexible actuator (4) includes a movable unit (6) capable of bending, a pump unit (8) integrally connected to the movable unit (6) at an end thereof, for controlling transfer of work fluid sealed in the movable unit, and a pump driver (10) for controlling driving of the pump unit (8). The movable unit (6) includes two long and narrow tubes (12) and (14) each having an accordion wall, a core (16) maintaining the tubes in parallel to each other, and a thin-film-bag-like cover (18) covering the tubes (12) and (14). The tube (12), which comes into contact with a subject (B), is made of a soft material. The movable unit (6) provided with the tube (12) softly comes into contact with the subject at a plurality of plane-contact portions of action of the movable unit (6), thereby holding the subject without scratching the same.

**Correspondence Address:**  
**Oliff & Berridge**  
**PO Box 19928**  
**Alexandria, VA 22320 (US)**

21) Appl. No.: 09/979,177

22) PCT Filed: Mar. 2, 2001

86) PCT No.: PCT/JP01/01660

30) **Foreign Application Priority Data**  
Mar. 28, 2000 (JP) ..... 2000-089581

**Publication Classification**

51) Int. Cl.<sup>7</sup> ..... F16D 31/00

743 x 1,091 | 288, 4 | 809, 71 | 521 x 67

C:\Users\dauda\Desktop\Search Light tagger\espc\New folder\US20020157388.pdf

**ZONES**

Zone 2

**Type** **Name** **Title** **Patent ID** **Keywords**

☐ US20020157388 PUMP-INTEGRATED FLEXIBLE ACTUATOR US 20020157388A1

**Add document**

**Zone Name** Zone 2

**Capture** ☐ Text ☒ Barcode

**Barcode Type** Aztec,Codabar,Code 39,Code 93,Cod

**Where** Barcode matches any pattern

**Advanced Settings**

**From** All pages

**SharePoint column to map to** Patent ID

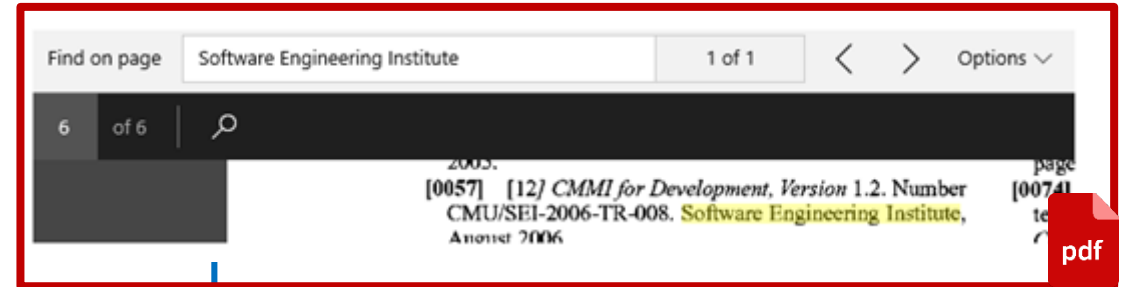
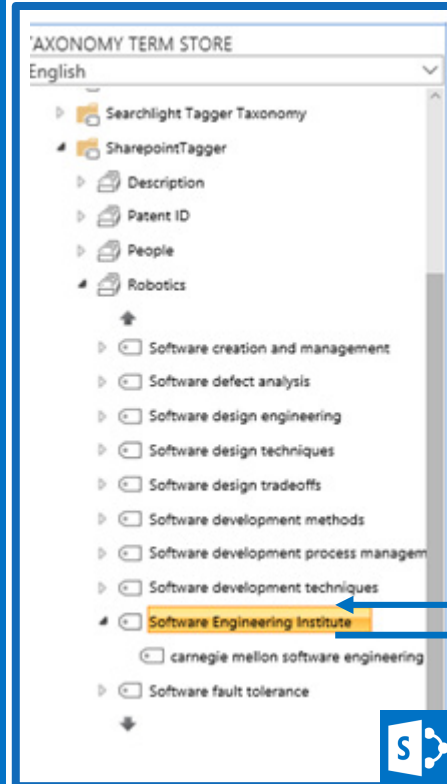
**If SharePoint column already has value(s)** Append

**Tag Limit** 0

**Test Extraction**

**Save** **Cancel**

# Taxonomy Matching

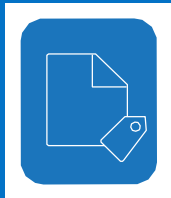


Searchlight matches the content of your documents to a term store in SharePoint to see if any terms appear in the document

If any hits are found, Searchlight will then add them as values to the Managed Metadata columns

<input type="checkbox"/>	Type	Name	Title	Patent ID ↑	Keywords
		US20110196798	PROJECT MANAGEMENT ROBOT METHOD AND SYSTEM	US 20110196798A1	Software Engineering Institute; management; Documentation; industry standard; constraint; Measurement; quality management; Technologies; markup language; Organizations; empirical study; Constraints; industry standards; Barry W. Boehm; software architecture
Add document					

# PDF Metadata Extraction



Custom Properties

Name: PatentID

Value: US007590680B2

Name	Value
PatentID	US007590680B2

pdf

Searchlight can map PDF Metadata (Default/Custom) to SharePoint Metadata columns.

PDF Metadata	SharePoint column to map to	If SharePoint column already has value
Title	Title	Don't update
PatentID	Patent ID	Append

Aquaforest Searchlight

Type	Name	Title	Patent ID	Keywords
Final Form			US20020157388	
FIRA2015CFP0311			US007590680B2	

SharePoint

# PDF Forms



Home Tools FIRA2015CFP0311.... Patent Form.pdf

1 / 3

**PETITION TO ACCEPT UNIN**

Patent Number	Issue Date (YYYY-MM-DD)
US200201573	2017-10-18

pdf

Searchlight can map PDF form fields to SharePoint Metadata columns.

Load Form Fields

PDF form field	SharePoint column to map to	If SharePoint column already has value(s)
Patent_ID	Patent ID	Overwrite

Aquaforest Searchlight

ESPC17 Presentation EDIT LINKS

After Library

Search...

Type	Name	Title	Patent ID	Keywords
Final Form	Final Form		US20020157388	
FIRA2015CFP0311	FIRA2015CFP0311		US007590680B2	

# Image PDFs



## SHAREPOINT AUDIT RESULTS

### FILE SEARCHABILITY



## STATISTICS

### PDF Documents

Total PDF Documents:	2,421
Image-only PDFs:	80 (3.3 %)
Partially Searchable PDFs:	11 (0.5 %)
Fully Searchable PDFs:	2,300 (96.2 %)
Error PDF Documents:	30
Total PDF Pages:	18,027
Image-only Pages:	3,327 (18.5 %)
Fully Searchable Pages:	14,700 (81.5 %)

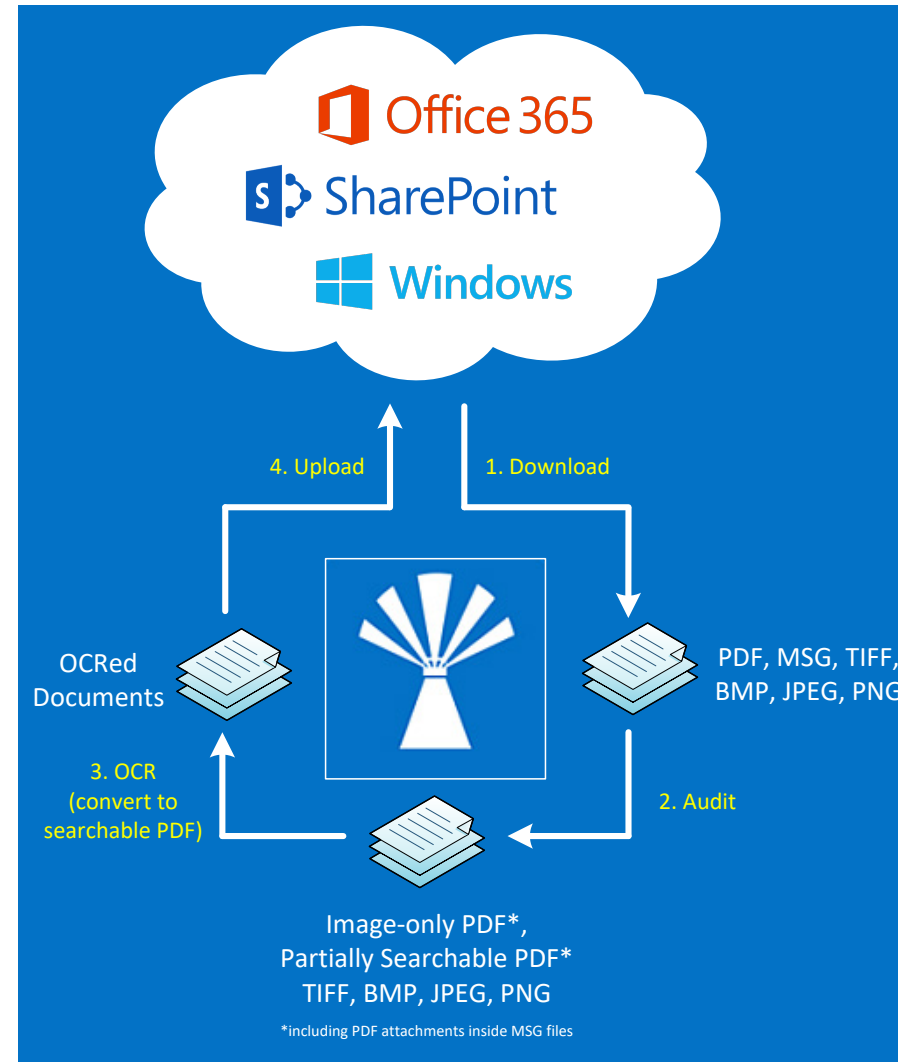
### TIFF Documents

Total TIFF Documents:	10,956
Error TIFF Documents:	83
Total TIFF Pages:	21,620

### Library Totals

Total Documents:	13,377
Total Error Documents:	113
Total Pages:	39,647
Total Searchable Pages:	14,700 (37.1 %)

# Image PDF Processing





# Searchable PDF Benefits



Text Searchable PDF

- 🔍 SharePoint Search
- ☰ Metadata Tagging
- 🔒 DLP Policies
- 📄 Content Reuse

Ask us for  
More Information

[info@aquaforest.com](mailto:info@aquaforest.com)