

# CONTENTCRAWLER

# Make All Your Documents Searchable

Improve productivity and compliance with contentCrawler's advanced Optical Character Recognition (OCR) and compression capabilities.



contentCrawler is an OCR and document compression solution that identifies non-searchable documents in a repository and converts them into text-searchable PDFs of manageable sizes – all while running 24/7 without any intervention.

# Why contentCrawler?

More than 30% of documents in any content repository are image-based and unsearchable. These invisible documents hinder the ability of legal professionals to find businesscritical documents and make it impossible for IT teams to properly manage and govern them.

contentCrawler runs 24/7 without any intervention and delivers detailed reports on processing statistics and errors, allowing IT teams to automate OCR and compression tasks, save on cloud-storage costs, and empower legal teams to be more productive. "contentCrawler is a no-brainer, which is very rare. Any technical concerns raised by users was overshadowed by the software's functionality, and the workflow it makes possible."

Kenneth Schultz CIO, Becker & Poliakoff



### **Smart Detection**

Leverage smart technology that finds and assesses image-based documents.



### 24/7 Automation

Make every document 100% searchable with an automated solution that runs 24/7.



## **Enhanced Compliance**

Ensure documents are searchable to comply with client and regulatory requirements.



#### Solution Overview contentCrawler

	Save All All	OCR All Search Doc # J Report Filter	Refresh					or Review		Re-Attempt Documents
-		neportrise	Live View Report	1 14	port Expe	n	maia 1			Exceptions
che		Nways active Every day	Document ID	Type	Paner	Original Size	OCR	Same	Status	Last Updated
		EMP_UPG_MSG_POF_PNG_TIFF	Matters 587484 v2.0	PDF	10	425 KB		Save	Completed	23/07/2016 10:02 PM
6.01	ch (	DMP, JPG, MOG, JPDF, PNG, TIFF	Matters 588649 v1.0	MSG	3			Save	Completed	23/07/2016 10:02 PM
			Matters 587480 v2.0	POF	10			Save	Completed	23/07/2016 10:02 PM
e	ration	OCR, Save as New Version	Matters 588653 v1.0	MSG	12	1.7 MB	1	Save	Completed	23/07/2016 10:02 PM
à	serch & /	lasesa	Matters 588656 v1.0	MSG	7	97 KB	1	Save	Completed	23/07/2016 10:02 PM
		378417 of 378417 documents assessed	Matters 588654 v1.0	MSG	4	74 KB	1	Save	Completed	23/07/2016 10:02 PM
è	2	(2443478 pages assessed)	Matters 588657 v1.0	MSG	3	51 KB	1	Save	Completed	23/07/2016 10:01 PM
	V	190404 documents require processing	Matters 588661 v1.0	MSG	3	13.69 MB	1	Save	Completed	23/07/2016 10:01 PM
		[123.7 G8]	Matters 588658 v1.0	MSG	3	48 KB	~	Save	Completed	23/07/2016 10:01 PM
		OCR - 190404 documents (1414527 pages)	Matters 588662 v1.0	MSG	3	62 KB	1	Save	Completed	23/07/2016 10:01 PM
		0 exception(s)	Matters 588668 v1.0	MSG	4	S5 KB	~	Save	Completed	23/07/2016 10:01 PM
			Matters 588670 v1.0	MSG	6	1.07 MB	~	Save	Completed	2307/2016 10:01 PM
		Held for Beview No Documents	Matters 588665 v1.0	MSG	3	62 KB	1	Seve	Completed	23/07/2016 10:01 PM
			Matters 588682 v1.0	PDF	8	304 KB	~	Seve	Completed	23/07/2016 10:01 PM
	ocess.		Matters 588669 v 1.0	PDF	1	38 KB	~	Save	Completed	23/07/2016 10:01 PM
		190404 of 190404 documents processed	Matters 588697 v1.0	PDF	22	1.87 MB	~	Save	Completed	23/07/2016 10:01 PM
	~	OCR'd - 190404 of 190404 documents	Matters 588687 v1.0	POF	3	218 KB	~	Save	Completed	23/07/2016 10:00 PM
	•	(1414527 of 1414527 pages)	Matters 588698 v1.0	POF	18	2.06 MB	-	Seve	Completed	23/07/2016 10:00 PM
	. 7	0 secession(a)	Matters 588690 v1.0	PDF	8	410 KB	~	Seve	Completed	23/07/2016 10:00 PM
		Hold for Benjew No Documents	Matters 588688 v 1.0	PDF	3	203 KB	~	Save	Completed	23/07/2016 10:00 PM
			Matters 588689 v1.0	PDF	1	33 KB	~	Save	Completed	23/07/2016 9:59 PM
	ne.		Matters 588691 v1.0	POF	2	101 KB	~	Save	Completed	23/07/2016 9:59 PM
			Matters 588692 v1.0	POF	6	244 KB	-	Seve	Completed	23/07/2016 9:59 PM
	-	190404 of 190404 documents saved	Matters 588693 v1.0	PDF	4	238 KB	~	Save	Completed	23/07/2016 9:59 PM
	()	0 exception(a)	Matters 588694 v1.0	MSG	3	56 KB	~	Save	Completed	23/07/2016 9:59 PM
	V		Matters 588695 v1.0	POF	2	83 KB	~	Save	Completed	23/07/2016 9:59 PM
			Matters 588696 v1.0	POF	1	50 KB	~	Save	Completed	23/07/2016 9:59 PM

"contentCrawler jus sits in the background and does the job of OCRing for us whatever the source - i.e., scanned from a device, downloaded from an external portal or sent as an attachment by email."

**Christel Aguila** Head of IT, Winckworth Sherwood

# **Key Features**

### Convert to Searchable PDFs

Find image-based documents and non-searchable PDFs and have them automatically converted to searchable PDFs.

### Reduce File Size

 $\overline{}$ 

The Compression Module reduces file size, saving on data storage or migration costs without affecting document quality.

#### Dual Processing Mode

Process legacy documents that come in through scanning or mergers and acquisitions, and any new files created - simultaneously.

### 24/7 Automated Processing

The solution processes and compresses documents 24/7, running in the backend. Administrators can set it and forget it, while staff add and profile documents as usual.

### Dashboard Reporting

Use the Administration Console Dashboard to view progress on OCR and compression, see the number and percentage of documents processed, etc.

### Integrations

 $\overline{\square}$ 

contentCrawler finds and processes documents in document management systems, such as iManage and eDocs, and your file system.

For a closer look, visit <u>www.litera.com</u> or <u>request a demo</u>.