OCR TECHNOLOGY

Ensuring all your content is searchable!

By, Ray Zwiefelhofer
President, World Software Corporation

Many legal accounts I've spoken to in the past year did not understand the merits or concept behind OCR (Optical Character Recognition) technology and how they can benefit them. I thought I'd share some basics to help you to better understand what OCR technology can do for your firm and why you need it.

On a quest to become a more paperless and more mobile office, many law firms are using scanners more than ever to move paper hard copies into document management systems (DMS) or file shares to achieve the truly centralized 'electronic' matter file. Scanners and associated technology have been around for a decade or more and are increasingly straightforward with easier methods to help with the less-paper office workflow.

It is important to note that 'typically' when a scanner or a multifunction device scans a piece of paper it creates either a TIFF or PDF image file. Without the help of OCR technology, these files are non-searchable. Your firm may have a high-end document management solution with great built-in search features, but those features are worthless if many image documents are seen as pictures with no text and are non-searchable due to the lack of OCR. Most scanning solutions offer OCR as an add-in and not part of the base purchase.

Studies performed indicate that more than 20% of a firm's content repository (i.e., documents, email, scanned images) are NOT searchable. This means that one in five documents will not be returned in search results, but remain buried somewhere in the document management system. Non-searchable files usually consist of the following:

• CDs and DVDs from outside of your firm
• Attachments coming in through your email
• Paper scanned at your firm's multifunction devices

There are several methods to ensure that your all of your content is searchable depending upon the situation.

• Desktop. Most native apps such as Microsoft Office are searchable by default. However, there are applications from Adobe, Nuance and DocsCorp that allow searchable (OCR) PDF documents from the desktop. Such applications can of course be time-consuming and tie up your desktop computer.

• Personal Scanner - Personal scanners such as Fujitsu ScanSnap have built-in OCR technology that enable small jobs to be scanned from your desktop with OCR and saved to your document management system to make them fully searchable. Keep in mind that this is an option that you must choose. It takes a few seconds longer to process each document so I've seen many users disable this feature in favor of a quicker scan job. As a result, none of the scans are fully text-searchable by your powerful DMS search engine.

• Walk-up copier or MFD. Your large multi-function device does a great job at scanning large volumes of documents and is a good choice for mid-level volumes. Normally OCR technology needs to be added on through an additional purchase. Check with your copier vendor to ensure that you have this. One downside for larger jobs is the time it takes to process the job with OCR is significantly longer than without OCR. In one firm I recently visited, users consistently scanned 100+ page jobs and the wait at the copier was much too long.

• Server-based crawler technology. More recently, server-based solutions have been brought to market to resolve all of the above situations. A server-based solution typically resides on a dedicated server and runs all of the time to monitor your document repository. When installed it will scan through your old documents and process them to make them searchable. It OCRs every new scan you save to the system to make your full document repository searchable. This server-based solution has in the past been an expensive add-on in the tens of thousands of dollars. Recent market competition and other items have made this solution very affordable for even the smallest firm.

Now that you are aware of the various methods you can use for OCR, please consider the top reasons you should implement a proper OCR strategy at your firm:

• Reduce non-compliance risks – Failure to produce documents can have an impact on regulatory compliance and exposes an organization to unnecessary risk.

• Increase organizational productivity – OCR technology reduces productivity losses and downtime by finding misfiled documents or skimming through them to determine context.

• Knowledge repository – Chances are that your firm has contracts, briefs, etc. that go back for years. If you make legacy files fully available for your search engine you will expand your knowledge management and reduce time.

• Conflicts/new case intake – While this shouldn't be your only search, having knowledge that every document and email in your repository has been searched for a particular company or person's name could add extra comfort and accountability.

In order to reap the benefits of OCR technology, begin by examining all the content you receive and store in your document management system including email, client disks and scanned paper documents. Determine which files have had OCR technology applied to them. Next, consider the costs of implementing a reliable OCR strategy. Keep in mind that vendors who provide OCR solutions normally pay royalties to companies such as ABBYY and Nuance, creators of OCR technology. That cost is usually based on number of devices the technology is used on so you may want to not enable OCR on every device you have. To reduce the risk of paying for OCR on devices that do not need it, I'd recommend using a server-based OCR strategy. There are server-based solutions available today that are very affordable even for smaller firms. For a reasonable price you can run a server-based OCR solution all of the time to ensure that all documents, old and new, are fully text-searchable.

Mr. Zwiefelhofer has over twenty five years' product solution experience within the legal technology market with expertise in AMLAW 250 and Fortune 500. He was a President, CEO and CIO at several software solutions startups and the CTO at a Fortune 500 company. Those companies include Bowne, Imagineer, Equitrac and Diebold. Prior to joining World Software, Ray was the Founder and CEO for nQueue, a global cost recovery company.