Document Imaging Report

Business Trends on Converting Paper Processes to Electronic Format

4003 Wood Street ● Erie, PA 16509 ● PH (814) 866-2247 ● http://www.documentimagingreport.com

September 28, 2012

Kodak DI Presents New Software Face at Conference

LAS VEGAS—**Kodak's** Global Directions conference was certainly not Breakaway [the former name of Kodak's annual reseller conference] version 2.0. No, it was more like a complete reboot. Held at the J.W. Marriott resort in Summerlin, approximately 200 attendees were treated to some solid networking and a variety of presentations about trends in the document and information management market. One of those trends is Kodak Document Imaging (DI) transitioning its business to include more software and professional services.

"As we've said from the beginning, we saw Global Directions as more of an industry event that we are sponsoring, than a Kodak event," said Dolores Kruchten, who recently assumed the title of president of Kodak DI (while maintaining her role as VP at Kodak Corporate). "We were careful not to come off as too Kodak-centric. We offered a diverse agenda. We wanted attendees to look at different ways that people solve information management problems and draw their own conclusions."

Indeed, only about a third of the sessions featured discussion on Kodak products. None focused on the production scanners which Kodak DI has been best known for since its transition from a film to a digital focus more than 15 years ago. "We are continuing to expand our focus," explained Kruchten. "We're not forgetting about scanners. They provide us with a strong foundation, and we still spend more on hardware development than we do on software.

"But, with this event, we really wanted to shine the spotlight on our software products. Historically, our software has been mainly for enabling our scanners. It was all about making our scanners more effective.

"While scanning paper is still important to us, today there is a whole lot of information coming from other sources as well. Our new software focus addresses this. We realize that developing and delivering information management software is very different from marketing hardware. We didn't want to throw it all together.

"Especially when a new product line is in its early stages, it needs to have the right focus if it is going to succeed. That was one problem Kodak corporate had as it tried to transition to the digital market. It didn't give digital enough focus. We felt that with this event, we gave our new software products the right level of focus, and the attendees ample education opportunities."

At Global Directions, Kodak spotlighted two new software products. One was Info Activate, a capture solution for SharePoint that was announced a few weeks ago. The second was Info Insight, an IDR (intelligent document recognition) auto-classification and extraction product, which was previewed at Global Directions. Here's some more detail on each product.

Simplifying SP capture

Info Activate is designed to ensure that documents being stored in SharePoint contain the proper meta data. "Info Activate is an important tool for governance of SharePoint files," said Michael Frawley, CEO of **Edge Digital Group**, a Kodak reseller that has helped beta test the product. "Too often, SharePoint ends up acting as a giant file server, and people don't know how to find and open the appropriate libraries for specific documents. And, they don't know what index fields to enter.

"Info Activate lives and breathes inside of SharePoint sites. It forces users to enter the right meta data before they can save a file."

It also takes users to the appropriate library through a single click on an icon that represents a specific document type. "It's an amazing tool for getting scanned files into SharePoint, particularly if you have remote users," said Frawley. "It's truly one-click scanning.

"One of the biggest challenges with distributed scanning is that the people doing capture are not scanning technologists. A lot of times, you'll get images that aren't the proper resolution, or that don't have the right rotation and cropping applied. Info Activate enables an administrator to lock down those settings."

Info Activate is a Silverlight-based application loaded into SharePoint. It works with SharePoint 2010. It can run in multiple browser types and connect to most TWAIN-driven scanners. Documents can also be uploaded into it from other sources, such as a desktop file system.

When users who are logged into SharePoint launch Info Activate, they receive a customized interface. They are presented one or multiple buttons for scanning. These buttons are labeled with the document types the user has permission to scan. An administrator sets up the scanning parameters for each doc type—such as resolution and compression. Image processing can be done through the TWAIN driver and complemented through additional options on the Info Activate server.

Batches of documents can be captured and split through separator sheets or manually. Bar codes can also be read for capturing index information. Meta data fields will be automatically assigned to a captured document based on the library the document is being captured to.

For instance, if the meta data fields for an invoices library are "vendor," "date," "item number," and "total fields," these fields will be assigned to any documents captured with the invoices library as their destination. The destination of a captured document is pre-set by the administrator.

A screen showing the meta data fields will pop up before a captured document can be saved. The data can be manually keyed or a user can select the appropriate text on the document and drag it over to complete the field. Indexing can be completed at the time of scanning or in a post scan process (which enables indexing to be done at different sites than scanning).

Documents can be saved as TIFFs or PDFs. They can also be exported from SharePoint, along with meta data in an XML file, for archiving in another ECM system. In a demo at Global Directions, Kodak showed Info Activate integrated with **Nintex** workflow to enable an approval process. Kodak also showed its Document Viewer for SharePoint, which it introduced last summer [*see DIR 5/20/11*].

Pricing for Info Activate is based on a server license, plus a specified number of concurrent users. A development server license is also available.

There will also be a light version called Capture Office available through the **Microsoft** Office 365 Marketplace. This is a SaaS implementation designed to work with Microsoft's hosted version of SharePoint. A more broad-based SaaS implementation of Info Activate and a mobile app are on the way.

Document Imaging Report
Business Trends On Converting Paper Processes To Electronic Format

DIR is the leading executive report on managing documents for e-business. Areas we cover include:
1. Document Capture
2. Image Processing
3. Forms Processing/OCR/ICR
4. Enterprise Content Management
5. Records Management
6. Document Output
7. Storage

DIR brings you the inside story behind the deals and decisions that affect your business.

Vol. 22, No. 17

Editor: Ralph Gammon
4003 Wood Street
Erie, PA 16509
PH (814) 866-2247
FX (412) 291-1352
ralphg@documentimagingreport.com

Managing Editor:
Rick Morgan
PH (814) 866-1146
rickm@scandcr.com

DIR is published 23x per year, on the 1st & 3rd Fridays of the month, by:
RMG Enterprises, Inc.
4003 Wood Street
Erie, PA 16509
PH (814) 218-6017
http://www.documentimagingreport.com

Subscriptions: $597 (electronic) or $670 (paper) per year.

For more information: http://tinyurl.com/KodakIA; http://graphics.kodak.com/DocImaging/US/en/About_Us/News/2012/120830a.htm

Kodak introduces advanced capture

Info Insight is a true multi-channel capture application that can be utilized on scanned documents, as well as electronic documents like e-mails, text messages, and even social media communication. It is based on technology licensed from the German ISV **ITyX**, which markets its document understanding software to both the digital mailroom and call center markets. ITyX has implementations with several large companies in Europe.

"We have definitely seen a need for intelligent capture in the market," said Robert Bijster, Kodak DI's worldwide director for software marketing and commercialization. "We wanted to have a way to go beyond capturing paper documents. Yes, paper is where our heritage lies, but our customers are now dealing with multiple types of input.

"In addition, we wanted to be able to handle unstructured content. Many people, including Kodak, have figured out good ways to process structured content. We have tens of thousands of implementations of Kodak Capture Pro worldwide that can handle document capture up to a certain level. But, we think the future lies in a solution like Info Insight that can handle structured, semi-structured, and unstructured documents."

Info Insight is being marketed separately from Kodak Capture Pro, which can potentially be configured on the front end for capturing paper batches, or on the back-end for providing release into third-party applications. Some of the configuration details are still being worked out and Info Insight is not scheduled to be released until early 2013. It will be offered as both a cloud and on-premise solution.

"We think the cloud offering will enable us to serve the mid-market, which has historically been very underserved by IDR," said Bijster. "The cloud really helps reduce customers' barrier to entry. We will develop some base, repeatable applications that we will make available through the SaaS model. But, there will also be opportunity for our partners to customize cloud solutions leveraging their own expertise."

ITyX's technology was originally developed to process e-mails, which have notoriously little structure to them. We saw a demo in which a very poorly written e-mail, purposely littered with bad misspellings, was run against a fictional database of several thousand customer names. Within a fraction of a second, Info Insight was able to determine which customers the e-mail was referring to.

"We can basically push content through the database," explained Süleyman Arayan, founder and CEO of ITyX. He compared the process to a strainer which holds back only the information that best matches the content of the database. "But what if you receive a communication for which you don't even know what database to compare it to? What if you receive a contract from a new customer? This is when we apply our artificial intelligence and self-learning."

A learn-by-example approach can be used to basically teach Info Insight how to extract data from a variety of documents types. This same approach can also be used to train the system on auto-classification. Rules can be applied on more structured document types.

An evolving business

Kodak DI realizes that as it gets into more complex products, such as the IDR, it needs to evolve its channel and services. "Our channel is going to have to expand," said Kruchten. "Historically, we've been heavily focused on a channel that manages paper processes. But, that's not enough anymore. Our partners need to look at the entire flow of information.

"Some of our software resellers will be scanner resellers. But, others will come from different areas—like SharePoint integration."

Of the approximately 120 VAR/systems integrator representatives at Global Directions, Jackie Horn, Kodak DI's worldwide director of marketing, said a good deal were not current Kodak resellers. "We had a good mix of attendees," she said. "Some have a long history in document management and are interested into integrating it with other platforms. But, we also reached out to new organizations, like those with an interest in collaboration."

To address its evolving market, Kodak DI is building out its professional services capabilities. Some of this started before it was announced DI was for sale, when Kruchten was serving as Kodak corporate's GM of enterprise services and solutions [*see DIR 1/20/12*].

"In some areas, we're are still figuring out exactly who will stay with Kodak corporate and who will go with DI," said Kruchten. "But, the bottom line is that DI has to build out its technical support capabilities to support our emerging software business. We need to provide customization and support for the channel. We had to do the same thing when we first start selling document scanners. It took 10 years for scanners to become a catalogue number sale."

Kruchten indicated that some of the new professional services personnel could be transitioned from DI's Service organization, which to date has primarily focused on hardware maintenance. "Hardware service and software services are separate businesses," she said. "But, our hardware service revenue is relatively flat to declining slowly [due to the market moving more toward distributed, lower priced scanners, which are often treated as disposable items]. We have a lot of talented people working for us in Service, and we'd like to move some of them to professional services."

Looking forward to a bright future

In her keynote, Kruchten indicated that there have been plenty of interested potential buyers for DI, and it seems that almost everyone in the organization is looking forward to the pending sale, which is supposed to close during the first half of 2013. "By the end of the year, we should have a completely refreshed hardware line," Kruchten told *DIR*. "We have the best field service organization in the market. And, software represents a whole new avenue that we are excited about. We are looking forward to new ownership to help boost our investment in going to market with our new offerings."

For more information:
http://graphics.kodak.com/docimaging/us/en/index.htm

NSi Announces Ambitious Mobile App

In addition to more intelligence being introduced into capture software, one of the major themes to emerge at the **Harvey Spencer Associates'** (HSA) recent 2012 Capture Conference was the need for capture software to serve as an input channel for more than just paper. Indeed, Spencer cited e-mail, Web sites, social media, tablets, and cell phones as possible avenues of data entry.

Notable Solutions, Inc. (**NSi**) hasn't quite gone that wide, but its new NSi Mobile is certainly a multi-channel capture app. "We play in what Harvey defines as the transactional capture segment of the software market," said Mike Morper, VP of marketing for Rockville, MD-based NSi. "We believe our segment encompasses far more than paper capture.

"It was nice to see that theme emerge at his conference. We want to capture any asset an organization needs to drive a process. This could be paper, but it could also be photo or an e-form, and it could come from anywhere."

When Morper says "anywhere," he means it. NSi is best known for its technology for capturing documents with MFPs. NSi Mobile moves the company's capture technology onto smartphones and tablets.

"Our goal has always been to bring a personalized user experience to the front panel of an MFP device," said Morper. "With AutoStore 6.0 [the brand name for NSi's flagship application], we really expanded the scope of the information that could be captured. We added support for electronic documents and XML data streams. We believe our experience uniquely positions us to help professionals utilize their mobile devices for transactional capture.

"Fundamentally, our goal is to capture information that comes in from anywhere and needs to be coded and delivered to a business process. For the past half year, we've been getting inquiries from our customer base asking us to help them be more productive when they are away from their desks. This doesn't mean they have to be out in the field. It could just be people moving around the office or sitting in a meeting and multi-tasking with their tablet or smartphone."

NSi Mobile consists of a server piece that connects to AutoStore 6, as well as an app that can be downloaded onto both Android and iOS devices. "Our connection with AutoStore is one of our value propositions," said Morper. "It enables NSi Mobile to take advantage of AutoStore's security features as well as its large number of connections to back-end systems.

"One of our differentiators has always been that AutoStore is a highly secure capture application. We utilize Active Directory for authentication. The NSi Mobile server runs in a DMZ on the network and talks to the AutoStore server, which is behind the user's firewall.

"AutoStore can be utilized to deliver documents into 40 ECM systems, as well as several line of business applications. Through AutoStore, we can also pre-populate data fields related to items captured with NSi Mobile. AutoStore can act as a broker between line of business systems on the back-end and NSi Mobile on the front."

Broad set of use cases

The NSi Mobile app is a relatively small (approximately 2 MB) and will be available next month through Google Play or the Apple App Store. The server piece is available to anyone who has purchased AutoStore 6's Web Capture option [*see* *DIR 1/20/12*].

There are four main use cases being targeted with the initial version of NSi Mobile:

■ **document and photo capture: "**For documents, the most obvious example is someone on the road capturing trailing documents—like a loan officer capturing tax forms and pay stubs at an applicant's house," said Morper. "When a user logs in, AutoStore knows what rights they have and could present the officer with an e-form for a loan application, for example. In addition, AutoStore would ask him for pictures of certain documents.

"Those documents can be captured through the smartphone camera and submitted with the e-forms data. Our internal edict is that anything being done on an MFP panel should take less than 10 seconds. We extend that to our mobile app."

Basically, a user takes a picture of a document, gets a preview, and then chooses whether to submit the image or reshoot it. "We don't think there needs to be any image processing for this type of process," said Morper. "It's also the most basic concept for capture with a mobile device, and I think it will be our least adopted use case.

"We think taking a photograph to contribute to a business process is more relevant, for example. We have a customer in Europe that manages wind turbines, and they often need to provide damage assessments. Currently, they use digital cameras, which they have to attach to a laptop to submit the photograph and accompanying data to a workflow. With NSi Mobile, they should be able to transition the whole process to a smartphone."

Morper said that some NSi customers have asked about video capture. "It's definitely a viable use case, but you have to be able to deal with some significant file sizes," he said.

In addition to photos and documents, NSi Mobile can be used to submit documents already on a mobile device into a business process. "Say you receive an invoice as an e-mail attachment and you want to submit it to your A/P workflow," he said. "You basically click on it and when the 'open in' option pops up, you choose NSi Mobile. You then select your invoices tab, which will present you with an e-form for entering the appropriate meta data.

"This works for any app that can present users with the 'open in' option. This includes on-line file storage apps. Almost any file that you can access from your mobile device can now be handed off through NSi Mobile to a pre-defined workflow. We think this is a real home run."

■ **Mobile e-forms:** Basically, this option uses the same interface and data look-ups that are used when submitting document images and photographs to an AutoStore workflow—but users are submitting straight data with no image. "We think this feature differentiates NSi Mobile from the capture apps of traditional document imaging competitors," said Morper.

■ **Mobile pull-print:** "Pull-print is basically printing to a queue and then having to authenticate at an MFP to actually print your documents," said Morper. "We've been offering this functionality for about a year. It helps maintain confidentiality of print jobs, as well as saves paper, because a lot of stuff output by network printers never gets picked up.

"With NSi Mobile, users can launch jobs queued in our SecurePrint server through their mobile devices. They can do this in one of two ways. They can use a GPS look-up to find nearby printers registered on their network. Or, they can use their smartphone to photograph a QR code on a sticker [which we provide them with] attached to the MFP they want to print from. This will trigger a process through which SecurePrint will go to that user's queue and print at that device. This is the fastest way on the market to do pull-printing."

■ **secure MyFiles access:** "MyFiles is basically a user's home folder that is kept on a network and managed through Active Directory," said Morper. "Many organizations are trying to get employees to stop storing files on services like Box and Dropbox and encourage MyFiles as an alternative. Through NSi Mobile, users can view their MyFiles documents on mobile devices. We believe we will start receiving feedback to enable access to other network folders, but we thought it made sense to start with users' home folders."

For more information:
www.nsiautostore.com/autostore/mobile

More from recent HSA Conference

Last issue we covered a lot on document analytics, as discussed at **Harvey Spencer Associates'** recent Capture Conference. But that was not the only topic covered at the event. Here's a sample of some other stuff we saw and heard:

■ In his review of capture software sales for 2011, Spencer noted that the line is blurring between what he has traditionally defined as batch image and batch transactional capture. "This is occurring as vendors who historically have focused on batch capture have embedded OCR and other automated recognition technologies in their applications," said Spencer. "Next year, I may blend these two segments together.

■ Spencer noted that in 2011, "enterprise"-level capture software sales (which make up more than half of total sales) increased by 6%, while sales to the SMB market (just a small percentage to start with) actually fell off by 4%. "I think this has to do with the increasing sophistication of capture technology," Spencer noted. "A lot of SMBs that are using imaging have already made the decision to buy a desktop application like PaperPort. They are now trying to decide if they should make an investment to move upstream."

■ Spencer continues to be very aggressive on his projections of sales related to mobile capture software. He predicted that by 2017, the ad hoc transaction capture segment, which envelopes mobile capture, will surpass batch transaction capture, which includes traditional forms processing and IDR applications.

■ *DIR* Editor Ralph Gammon hit on four of the five predictions he made at the 2011 Capture Conference. The only one he scored as wrong was "SharePoint emerges as ECM platform of choice for dedicated capture ISVs." However, Mike Alsup, senior VP of systems integrator **Gimmal**, thought I was being too hard on myself.

"I thought you understated the impact of SharePoint 2013 on the ECM market," said Alsup. "I think SharePoint and the cloud are hollowing out the ECM market. Most big companies seem to be on a path to adopt SharePoint as their portal, ECM, RM, and unstructured content management platform.

"They are starting with shared drives and collaboration applications. But, a company with 100,000 employees will have up to 50,000 SharePoint sites, and they can't wait to simplify their stack and reduce reliance on legacy ECM suites. From a big company perspective, it isn't just license and maintenance costs [working against legacy ECM products], it is idiosyncratic development tools, proprietary platforms to maintain at great expense, and users who just don't like them.

In most cases, SharePoint is winning from the bottom up with the enthusiastic support of the CIO. When you talk about cloud and Azure, it isn't clear to me why people develop in Azure except to expose their services to SharePoint users via Office 365."

■ Priscilla Emery of **e-Nterprise Advisors** gave a talk on applying records management to documents stored in the cloud. She noted that one of the biggest challenges is knowing the physical location of cloud servers storing documents. Many governments have regulations regarding certain types of documents leaving their jurisdiction. "Ultimately, end users are responsible for the location of their documents," she stressed.

■ Dr. Arif Esa, solution manager within line of business finance at **SAP**, discussed the new Travel Receipts Management application by **Open Text** that SAP is selling through its OEM partnership with Open Text. The application utilizes the automated capture technology Open Text acquired with Captaris ODT [*see DIR 9/12/08*] to automatically extract and categorize data from travel receipts. It's integrated directly with SAP's travel and expense management software.

■ Finally, offline we caught up with Dmitry Harchenko, business development director, for **STOIK Technology**. STOIK is a Moscow-based developer of image and video processing software. At Capture 2012, Harchenko showed us a document capture app for mobile phones, and discussed his company's SDK, which is aimed at ISVs that want to incorporate mobile capture in their apps.

"Our goal is to provide the last mile in applications utilizing mobile document capture," said Harchenko. "We provide technology that can capture and process an image on a smartphone, without an Internet connection. The user can then upload that image to the application, destination, or business process of their choice."

MDScan, the consumer version of STOIK's technology, is less than a 2 MB download. It offers features like auto-cropping, noise removal, thresholding, rotation, and brightness normalization. It will automatically apply all these fixes to a document in a few seconds, and there are also manual controls for making adjustments. There is a batch mode for multi-page documents and several different types of output options including black-and-white, no enhancement, low-light shot, business card, receipt, color document, white board, etc. Documents can be formatted as PDFs or JPEGs.

Image processing can be done immediately after capturing a document—to ensure immediate feedback, or delayed for a future time (spy mode).

"We don't apply any OCR," Harchenko explains.

"That is better done on the server. We really envision our product as the front-end to BPM and document management systems."

STOIK's capture technology is currently available for Android operating systems with an iOS version in beta testing. "From our vantage point, most of the corporate world is working with Android devices," said Harchenko. "We see it transitioning that way from Blackberries."

STOIK's SDK is currently available and is already being licensed by ISV's like **OfficeDrop**, which is utilizing the STOIK technology in a mobile app it uses to feed its cloud repository. "We already have several corporate licenses," said Harchenko.

For more information:
http://www.harveyspencer.com/documentcapture;
http://www.gimmal.com;
http://www.ecmscope.com/index.html;
http://www.sap.com/lines-of-business/finance/travel-receipts-management;
http://www.stoik.com/products/mobile/mdscan/

ISV Takes Fresh Look at Recognition

There is no question that ISVs are currently trying to go where no capture software has gone before—in terms of applying automatic recognition technology to documents. Over the past couple issues, we've covered topics like artificial intelligence, semantics, and advanced analytics, all designed to take applications beyond the capabilities of current capture. However, an appropriately named start-up out of Germantown, TN (near Memphis), may have beaten many established capture players to the mark.

Using a combination of advanced pattern recognition and computer vision, **BeyondRecognition** recently completed a project in which it successfully indexed 2.3 billion images, which originally contained no meta data. Yes, working as a contractor for an energy company, BeyondRecognition was presented with 27,000 CDs and DVDs full of images that covered a timeframe of roughly 90 years and were created in different locations across the world.

"There were no boundaries separating the scanned images," said John Martin, founder and CEO of BeyondRecognition. "The only thing we knew was that the documents on the discs in the front of each box were scanned before the documents on the discs in the back. Our client wanted to be able to mine the data on all these documents."

According to Martin, he looked at everything available for accomplishing the task at hand. "I looked at traditional OCR applications, but they were a bust—even with voting engines (which would just have made the process take five or six times longer)," he said. "Even if we could have applied full-text OCR, conventional search engines could not do the things we wanted. In addition, traditional relational databases couldn't support the millions of many-to-many relationships we had to set up."

To build the application that eventually became of cornerstone of BeyondRecognition, Martin applied a process he called "negative learning." "Basically, I started with no pre-conceived ideas about how automatic recognition was being done," he said. "Instead, I tried to take the position that if I were to build a recognition solution with tools available today, how would I approach it?"

The guts of the system

Martin started with the basic premise that computers are good at working with numbers. "Based on that, we were able to identify one of the key flaws of traditional OCR—it attempts to read characters like humans do," he said. "It goes left to right, top to bottom, first page to last.

"And it attempts to recognize each character individually. Think about that from a statistical standpoint. On each page, a single lowercase character might appear 40 to 50 times. So, on a million pages, that character could show up as many as 50 million times. Basically, with traditional OCR, you're giving software 50 million chances to get it wrong. Statistically, that means, it's certainly going to make at least one mistake."

BeyondRecognition puts each image through a process it calls "scraping." "We literally rip the images apart into glyphs," said Martin. "Those glyphs include not only the characters on a page, but also things like logos, staple holes, check boxes, signatures, and even specks of dirt. On average, we produce about 1,500 glyphs per page.

"We then run the glyphs through a normalization process before grouping them. The normalization involves accounting for orientation by rotating each glyph 720 times in half-degree increments. This way, direction doesn't matter, and neither does size. After it's normalized, if that glyph is at least 99% similar to other glyphs, they are placed in the same cluster."

What happens next is a bit confusing, but it basically involves identifying these clusters as sets of characters. This is accomplished at least partially by identifying the glyph in a cluster that most exactly resembles a known character and then plugging that character into a word that is checked against a global dictionary. There are also statistical formulas incorporated regarding how often a particular character should show up in a set of documents.

"We average the results of all that, and if it comes back above a 99% confidence rating that the glyph represents a specific character, we presume it to be true," said Martin. "Then, because our software tracks the location of each glyph it creates, we can identify all the glyphs in that particular cluster as being that particular character."

Of course, not every glyph comes back at a 99% confidence level. To account for this, Martin showed us a process called "Word QC." In the example, "lockbox" was not recognized as a valid word in the global dictionary, so it was highlighted. The statistics said it was one of several million suspect words (in a large set of documents). Merely confirming that "lockbox" was a valid word had a cascading effect that helped validate other glyphs as characters. The result was that with a single keystroke, 900,000 suspect words were eliminated.

BeyondRecognition has the ability to output searchable PDF files, as well as what it calls an "XPDF" file. "Basically, this is a cross-reference file, which includes a coordinate point for every word and numeric sequence pulled off a page," said Martin. "It's a great tool for redaction applications. We have a customer using it to redact expressions like Social Security and phone numbers. They can achieve a rate of 600,000 redactions per hour."

Because BeyondRecognition works with glyphs, it is able to handle multiple languages—even mixed within a single document set.

It also has the ability to do document clustering based on the layout and content of images. "This is a great tool for automatically routing files to the right process," said Martin. "We have a BPO that utilizes that element to help it allocate its resources more effectively.

Rules can be set up within BeyondRecognition's application for extracting specific data fields and tables. Extraction can be applied to structured and semi-structured documents. Rules can also be set up around the glyphs to eliminate background noise such as watermarks. Martin showed an example of image enhancement being applied to documents created through carbon-paper duplication.

Martin said the speed of BeyondRecognition's software depends on the number of CPUs being utilized. "A 30-core server can process up to a million pages per day," he said. "An 80-core can do 5 million, and a 160-core, about 10 million.

To date, BeyondRecognition has offered its technology solely as a service. "We've done several dozen projects," said Martin. "We're currently working on developing an appliance that can be run behind a customer's firewall."

Martin said that BeyondRecognition's technology belongs entirely to his company. "There are some patents around it, as well as some we're applying for," he said. "We have 15-16 man years worth of development in this."

For more info: http://www.beyondrecognition.net