

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 2, 2016

**DON'T TRY THIS AT HOME:
AUTOMATING CAPTURE FROM
UNSTRUCTURED HANDWRITTEN
FIELDS**

The automated data capture market is fairly mature. Users have been successfully applying character recognition in forms processing applications for more than 20 years. This works especially well on forms with well-structured fields and machine printed text. But, what about forms with handprint or handwritten text? And what if the fields aren't so well structured? What if they include open ended answers that can span multiple lines? Well, let's just say that to date, there hasn't been a whole lot of automation applied to capturing those documents.

Systems integrator **EDAC**, along with a couple of ISV partners, is hoping to change that. EDAC, which is based in Fredericksburg, VA, recently announced its new ReadScript Maestro module, which leverages technology from handwriting/handprint recognition specialists **Pasascript** and **Natural Intelligent Technologies (NITe)**. Pasascript, which is based in the Boulder, CO-area, has worked with EDAC for several years and supplies word recognition for ReadScript, a module developed by EDAC primarily for adding handwriting/handprint recognition to **IBM** Datacap forms processing applications. It's NITe, an Italian ISV, that adds the special sauce to ReadScript Maestro.

"ReadScript has always worked well in enough on fields where you might be asking a 'yes' or 'no' question or capturing some sort of structured information like a name, address, date, or identification number," said Randy Blevins, COO of EDAC. "In these instances,

CONTINUED ON PAGE 7

EDAC MAESTRO, FROM PAGE 1

we've also been able improve accuracy by taking advantage of validation sources like custom dictionaries, such as a list of drugs if it's a medical form, or a list of addresses from the U.S. postal dictionary. In the latter case, we can also use techniques like validating zip codes against cities and street addresses.

"The big challenge has always been handling unstructured fields, like you might find on a medical form, asking for a description of a condition. Similarly, on a loan application form, there might be a field asking for a job description—some people might give a one-word answer, but others can go on and on."

This is where Maestro comes in. "Maestro is able to look at this unstructured content and separate it into lines and single words," said Blevins. "It then passes the individual words on to ReadScript, which can recognize them and put them in order so they can be used as indexing values in a Datacap application."

At the core of Maestro is technology from NITe, which was founded in 2012 as a spin-off from the University of Salerno, located just south of Naples. According to NITe' Web site, "Our mission is to commercialize the technologies created and developed in the Natural Computational Laboratory. NITe provides innovative technologies and advanced software solutions for automatic processing of digital documents, especially handwritten. The proposed solutions are based on an innovative and patented technology developed in 20 years of research, for automatically reading cursive handwritten documents."

"We have been working with NITe for close to a year," said Blevins. "We have developed a lot of the IP in Maestro, but the base model IP comes from NITe."

EDAC is currently working with a retail chain on

implementing Maestro to process sales orders. “They receive handwritten forms that they want us to extract all the sales information from,” said Blevins. “This includes the name of the company placing the order, the PO number, and all the line items.”

Art Randles, EDAC’s director of integration services, said that every place there are handwritten forms represents an opportunity for Maestro. “Probably 95% of the time you know what you are getting in a field, but there are plenty of forms in multiple industries—government, financial services, healthcare— where there are fields for comments or that include room for descriptions,” he said. “An investigation form for a police department or government agency is one example.”

Another area where Maestro has been proven to be effective is capturing e-mail addresses. “A lot of organizations will ask for people to write their e-mail addresses on forms,” said Randles. “We can get decent accuracy if we can break down those addresses into components—the name, the @ sign, and the service provider. That enables us to utilize validation tools like a dictionary of names and a list of service providers. Handwritten e-mail addresses are very difficult to recognize accurately otherwise.”

We asked what kind of accuracy levels Maestro can achieve. “If you can’t read something with the human eye, you’re not going to be able to read it with any technology,” said Mark Gallagher, VP of sales for Parascript. “But, if the text is fairly legibly, the recognition accuracy I’ve seen with Maestro is about the same as you get in most handwritten recognition applications, with improvements through adding context and that.”

EDAC utilizes the Datacap framework to manage verification and correction of anything the falls below acceptable recognition confidence levels. “We continue to work to enhance the technology and in a few weeks we expect to be releasing the next version of Maestro,” said Blevins. “It will be able to look ahead and behind at segments of text if it has identified some text as a word that does not validate as a word.”

Gallagher concluded that EDAC represents a premier partner for Parascript. “We provide the underlying recognition technology, and EDAC provides the business case expertise, as well as additional technology necessary to satisfy the use case,” he said.

For more information: <http://bit.ly/EDACMaestro>;
<http://www.nitesrl.com>

RECENT TWEETS

For updates between newsletters, please follow us on Twitter @DIREditor. Here are some recent Tweets:

Aug 31: New North American-based CEO with #SaaS experience named [@TopImageSystems](#); [@TISMSchrader](#) stays on as president; <http://bit.ly/TISnewCEO>

Aug. 24: New version of [@LexmarkSW](#) Invoice Capture Service SaaS offering integrated with Microsoft Dynamics NAV 2016 <http://bit.ly/LexmarkICS>

Sept. 1: Two new desktop document scanners introduced by [@EpsonAmerica](#) <http://bit.ly/Epsonscanners>

Aug. 26: Ephesoft Awarded \$5M Advanced Capture Project by Leading Financial Services Company <http://bit.ly/Ephesoft5M>