ESKER ®
Document Process Automation

Getting the Value Out of Documents

Derby, UK – June 6th, 2011 – As companies seek to reduce the use of paper, document capture is only part of the answer, suggests Stephen Wynne-Jones, Technical Director at Esker

No matter what channel a document comes through and what type of document it is, the challenge is how to remove manual handling inefficiency and risk of errors. When the information from incoming business documents has been captured, it needs to be analysed, understood and managed. This means putting the information through a workflow for validation and eventually routing this information to a business application and/or a storage repository. The more you can automate this process from end to end, the more value you can extract from inbound business documents.

OCR technology has provided a functional means of making business documents machine-readable, and is most practical for structured documents, which have formats and layouts that do not vary. But human intervention may still be required to assure the quality of the captured data.

Today, Dynamic Document Capture technology offers additional automation intelligence for automated processing of semi-structured documents like sales orders and invoices, which contain consistent information in varying layouts and formats. Dynamic Document Capture improves recognition and accuracy, helping organisations eliminate traditional issues with managing different types of inbound business documents.

THE BUSINESS OF CAPTURE

Business cycles that have a direct impact on a company's performance involve processes that begin and end with a document. For example, the order-to-cash cycle starts with receiving a sales order and finishes with sending an invoice. Sales order documents reach a company via fax, email, mail (paper documents) or even by phone.

In the case of phone order processing, customer service staff enter orders directly into the customer order system, often after filling out a form. But if the document arrives by fax, email or mail, the customer service representative must manually enter the sales order information into the appropriate system. The process of managing such documents is inherently inefficient and labour-intensive, and carries considerable risk of errors and/or lost documents.

And the issue of documents in relation to business process efficiency improvement is here to stay - because the volume of documents received every day by companies is constantly increasing.

The types of documents a business typically receives fall into three general categories:

Structured documents

Semi-structured documents

Unstructured documents

To put things into perspective, today 20% of incoming documents are structured documents while 80% are semi-structured and unstructured.

STRUCTURED DOCUMENTS

Structured documents always have the same layout and an unchanging number of fields at a fixed position. For these types of documents, the goal is usually to extract data from a form and not necessarily keep the form. This data will then be migrated into a database for ERP, order entry, records, etc. The way to capture this information depends on the information. Such documents (health insurance claim forms, for example) are typically handled using OCR, ICR, OMR or barcode recognition. In this case, the best approach is to create a specific business rule for processing of each document type.

SEMI-STRUCTURED DOCUMENTS

Semi-structured documents have an unlimited number of variations, which makes the structured document approach inefficient. With semi-structured documents, data must be extracted from the document - no matter what its layout - for entry into an ERP application or other business system. Information can be automatically read and extracted from semi-structured documents such as sales orders, purchase orders and invoices, providing the ability to define a generic rule for each type of semi-structured document. This is possible because these documents always contain the same type of data introduced by keywords.

UNSTRUCTURED DOCUMENTS

With unstructured documents, the document itself is what's most important. You do not need to capture specific information from the page; you only need an image of the document so you can route and/or index it. Routing and indexing of such documents can be automated.

BEYOND IMAGING - THE LIMITATIONS OF OCR

Scanning documents can help to address data and document issues such as security, efficiency in data search and retrieval, document distribution and auditing. And in efforts to reduce paper handling costs, imaging has become a basic method to archive documents and data in image files.

To reach the next level of process efficiency, it is necessary to recognise the content of such files. Making data contained in documents available to other applications offers added value by replacing manual keystroke entry of entry into ERP systems. Moreover, automation of structured and semi-structured faxed or scanned documents relies heavily on successful image recognition. Skews and shifts, stretching and shrinking on images can lead to less than accurate recognition of defined areas.

Additional flexibility is offered by recognition that is not based on defined areas, but rather on the content of documents - keywords (located anywhere in the document) and fields to extract data and make intelligent assessments of document type and data required for processing. Compared with OCR alone, this type of technology offers vastly improved accuracy, combined with the ability to continuously "learn" how to handle different types of document formats automatically. The more a dynamic system like this is used, the more efficient it - and your organisation - becomes.

Recognition rates with OCR alone can be as low as 60%, depending on the quality of the electronic document and any additional technology needed to facilitate capture. OCR only reads an image and translates it into text. Beyond this, OCR does not "understand" content and how words relate to each other. In contrast, Dynamic Document Capture technology provides the ability to look for positions relative to key information. This additional business intelligence and logic ensures that accuracy of the capture does not only rely on OCR. This type of system is also able to run checks such as calculations and database lookups.

DYNAMIC DOCUMENT CAPTURE

Automation helps organisations raise operational efficiency by reducing document processing time and costs, whether the document is coming in from outside the company, flowing through the company or going out of the company to a customer or supplier. With Dynamic Document Capture you can speed up business processes and:

- Free your business from paper-based workflow
- Increase data entry accuracy
- Simplify document search and retrieval
- Enhance security of business data and documents.

Using business rules technology, Dynamic Document Capture allows intuitive rule customisation via a server-based graphical interface. Dynamic Document Capture incorporates OCR and ICR technologies to read information contained in image files whether they have been scanned or have arrived by fax or email. Additionally, Dynamic Document Capture uses logic, keywords, relative positions and business rules to allow relevant document data to be extracted. In a sales order scenario, the system automatically extracts the customer name and ship-to address, document date, number and total as well as line items such as quantity, description and amount. All information necessary to complete the sales order processing is captured directly from the document, regardless of its location within the document.

The ideal system comes with a set of predefined rules for processing for specific types of documents, such as sales orders. One rule extracts the information from inbound faxes or scanned documents. A validation form is used to validate extracted data. The output is an XML file which contains the extracted data and that is submitted to another rule for processing.

Dynamic Document Capture enables enterprises to avoid the cost-prohibitive task of defining a rule for each document variation, specifying the data you want to capture (data required for accounting purposes such as sales order number, invoice date, payment date, supplier references, totals, etc.) instead of specifying an area where it is located. Most of the time, data is introduced by keywords. For example, the total amount is introduced by the keyword Total. A generic rule captures data that is introduced by those keywords.

Because it is impossible to take into account the wide variety of different designs, Dynamic Document Capture lets you teach the generic rule to solve possible conflicts (for example, documents on which a keyword appears twice) or to improve system performance. Through a web interface, administrators and end users have the ability to improve recognition conditions or optimise document identification during document validation. With its intelligent business rules logic, Dynamic Document Capture is "free form" extraction that makes sense for semi-structured documents.

For more information, visit www.esker.co.uk.

North Europe Contact: Sam Townsend, Esker Ltd. – Tel: +44 (0)1332 548181, ☐ Email: samt@esker.co.uk
Corporate Contact: Renee Thomas, Esker, Inc. – Tel: 608.828.6140 ☐ Email: renee.thomas@esker.com
Investor Relations Contact: Emmanuel Olivier, Esker S.A. – Tel: 33 (0)4 72 83 46 46 ☐ Email: olivier@esker.fr

- ENDS -