

2009

Cabinet NG, Inc



DOCUMENT IMAGING-WHERE TO BEGIN

Table of Contents

Introduction	3
The Problem.....	3
Where to Begin	3
Document Preparation	3
Document Indexing.....	5
Ongoing Scanning of Daily Paper	5
Preparation and Indexing Examples	5
Example One	5
Example Two.....	6
Conclusion.....	7

Introduction

Perhaps the most asked question when considering document management solutions is 'How do I get my paper documents scanned quickly and efficiently?' Anyone contemplating document management or electronic archiving of paper based information must learn the options and determine what works best in their circumstances. These options must take into consideration the likely volumes of file cabinets or paper archives used in business and how to efficiently process the daily documents generated during normal business operations. The purpose of this paper is to introduce document imaging and guidelines for getting started with the scanning process. The entire process is commonly referred to as document capture or simply capture.

The Problem

Historical and current paper based business documents must be digitally captured to be used with an electronic document management system (EDMS). During the capture process, documents must be labeled with index information so they can be easily located once stored. Various document types may require different scanning parameters and index information depending on the type of document and how the document will be searched in the EDMS. Scanning thousands and perhaps millions of pages of paper requires a well thought out capture solution that is quick and efficient. Ongoing scanning of documents generated on a daily basis must also be addressed.

Where to Begin

There are many steps to consider before beginning the actual capture (scanning) process. The entire capture process can be broken down into five primary functions. This discussion will focus on steps 1 and 4. Steps 2,3, and 5 are more specific to selected capture products and will not be covered here.

1. Document preparation
2. Scanning
3. Quality Assurance
4. Indexing or Classification
5. Migration to the document management solution

Document Preparation

Document preparation involves a complete analysis of the current documents and the requirements for getting them ready to scan. This is the most critical and time consuming step of the capture process. Staples, paperclips, sticky notes, and general document cleanup are done during this step. A review of the documents to capture should answer the following questions:

1. What is the paper size of the documents? Are they all 8.5 x 11 or are some larger or smaller? Are they single or double sided?

This will help determine what size document feeder is required for the scanning hardware and if the scanner is simplex or duplex capable (ability to scan the front and back of the page). The scanner chosen should have the capability to scan the largest size document to the smallest size document without changing the scanning parameters. Specialized scanners should also be considered for unique requirements such as scanning insurance cards or large computer aided design (CAD) drawings.

2. What are the document characteristics? Are the documents all black and white or are some colored papers? Do the documents have highlighting, handwriting, pencil or ink on them?

These questions will help determine if you need enhanced image processing that is built in the scanner or a separate software add on to the scanner. Without enhanced image processing, colored paper, highlights and other document characteristics will cause poor imaging results resulting in unusable images (too dark, too light, etc).

3. Are the documents all the same type such as Invoices, PO's, Delivery Tickets, etc or are they mixed type like medical records, HR files, Client files, etc.?

This will help you determine how to classify the documents for the scanning step. If the documents are the same type they can most likely be scanned in batches with document categorization. If they are mixed type documents, they will need to be prepared for folder and document level categorization. This categorization is referred to as document separation. As the documents are scanned, the capture software can determine when a new document category is scanned or a new folder of documents is scanned based on the separator. Typical separators are blank pages, barcode pages or labels, patch codes, or fixed page counts.

4. Is there a requirement to search the text of a document once the images are stored? If so, what is the document quality? Are the documents laser, dot matrix, or fax printed?

Searching text within a scanned image requires an optical character recognition (OCR) process be performed on the image. If the document requires OCR, the document quality is very important in obtaining the desired results. OCR is never 100% accurate and the quality of the OCR results varies greatly depending on the condition of the original document, scan resolution, OCR recognition engine used, and content of the original document. OCR should be used in conjunction with a good document indexing scheme and should never be relied upon as the sole method of locating an electronic image in the document management software.

5. How many documents need scanning and how quickly must the job be done?

Determining the scanning volume will help determine suitable scanning hardware, number of staff required to complete the preparation and completion of the project in a specific timeframe. Scanners are rated in pages per minute so a calculation of time can be done based on the type of scanner selected. Time to prepare and index the documents will also need to be added to the scan time to make a final determination of project completion date.

Document Indexing

Document indexing is required to name the document effectively so it can be easily located once stored in electronic format. The index parameters can be determined by asking a simple question for each type of document to be stored. The question is ‘What information would I use to look up this document quickly?’ Indexing can be thought of in simple terms as the document title. Other index values can be added as keywords in addition to the document title. By using a combination of document title and keywords, a document can be easily located in the document management system. Document indexing also standardizes document information throughout the organization. Each person looking for a document knows what to look for because of the standardized naming convention employed.

Typical capture software will have index features available that allow for various index options. Some of the options may include: manual data entry, database lookup indexing, zone OCR indexing, barcode reading, and list box indexing. Each of these index types is beneficial in different situations to allow for the most efficient application of the document index values. Whatever method is used for indexing, it is important to find the most efficient way to apply the indexes for the documents being scanned so they can be processed as quickly as possible.

Ongoing Scanning of Daily Paper

The considerations mentioned above apply to daily scanning tasks as well. However, daily scanning may or may not require as much document preparation as back file scanning. If the daily scan volume is large (>500 pages), a high volume capture process will likely be the most efficient method to scan the documents. Daily scanning is also highly dependent on the source of the documents. If the documents are generated by in house systems, it may be possible to enhance the scanning process by having barcodes pre-printed on the documents. This will allow the capture software to classify and index the documents using automation thereby saving considerable preparation and index time.

For documents arriving from outside sources, manual key from entry classification and indexing will generally be applied. During the processing of these document types it is important to have a standardized naming scheme available. The use of naming templates and drop down lists of document descriptions ensure that documents are classified correctly and consistently, especially when multiple staff are involved in the scanning and filing process. This scanning process will also be best served with distributed capture and index or central capture and distributed index. The difference being whether desktop scanning is deployed or a central high volume scanner such as a multifunction device (MFP) is used to scan the documents. In either case, indexing software is deployed at the desktop level for classification and filing of the documents to the EDMS.

Preparation and Indexing Examples

Example One

Several file cabinets exist with invoices stored in them. The company would like to get all the old invoices scanned in quickly. After reviewing the invoices, it is found that all the invoices were printed on a laser printer and vary in length from 1 page to several pages. They are all 8.5 x 11 and single sided.

Some have highlighting on them and stamped information which contains handwritten data. The company would like to search the content of the invoices once stored and prefers PDF format. The company has 100,000 invoices to file and will hire a temporary high school student to do the scanning over the summer. The company would like to title the documents with the invoice date, invoice number, vendor number, purchase order number, and invoice amount.

Based on the invoice quality and properties defined, the company selected a scanner of 50 pages per minute scan rate with built in image processing to effectively capture the highlighted text and handwritten data in the stamp. The capture software selected produces searchable PDF output to handle the content search requirements. The index information was found to exist in the accounting database. This allowed the indexing to be setup for minimal user entry by using database lookup capability in the capture software. The only value the scan operator enters is the invoice number. All other index information is retrieved from the accounting database. To separate the invoices into individual invoices, a barcode label was placed on the first page of each invoice prior to scanning. The capture software created a new document each time a document separator barcode was read at scan time. With an estimated 40 work days to accomplish this task, the scanning and indexing of 2500 invoices per day was easily achieved.

Example Two

The human resource department of a corporation needs to scan employee folders into the document management software application for easy reference. Each employee has a file folder of documents in the HR department file system. Upon review of the documents, it is found that there are several different document types in each employee folder. The documents range in size from 8.5 x 11 and smaller and some are on colored paper. The document types can be classified into six defined categories and have specific document titles for each of the six categories. Each category can have more than one document title associated with it. Standard TIF file output format is sufficient and the content of the documents do not need searching. The estimated scan volume is 500 employee files with an average of 50 pages per file. The company selected a scanner of 30 pages per minute scan rate with built in image processing to effectively capture the colored paper documents. It allows scanning of paper from business card size to 8.5 x 14.

The preparation process includes removing staples and paperclips to separate the documents. A barcode is created for each of the six possible document types and the barcode is placed on the first page of each document type requiring titling. A folder separator is created on reusable sheets of paper. A folder separator is used for each employee folder. The folder separator will automatically create the required folder in the EDMS if it does not exist. Document indexing pick lists are setup for each of the six category document titles. This will allow the user to select the title from a drop down list with no data entry required. This also allowed multiple people to do scanning and ensured all documents would have the same title regardless of the scan operator. The company has limited staff time to dedicate to the scanning process. They allocated 2 months to scan the documents. This timeframe was achieved with the parameters selected.

Conclusion

The key to deploying a successful scanning solution begins with the planning stage. Understanding the document types to be scanned, hardware requirements, indexing methods available, and realistic timeframes to convert paper to digital images are all vital considerations. Without reviewing the physical documents and understanding the level of effort required to prepare the documents, the results may be disappointing. With careful planning and trial runs of the entire process, reasonable expectations can be set and met. Remember, if you need help, there are professionals available to assist you with the entire process.