# SimpleIndex

SimpleIndex provides one-click scanning and unattended processing for scan data capture including OCR, zonal OCR, and Barcode indexing capabilities.

How many clicks does it take to scan your documents? Many business scanning applications lack time-saving automation features like OCR and barcode recognition required for highly efficient document management systems.

SimpleIndex is an affordable fully-featured document capture solution that fills this gap. It is the fastest, most affordable way to organize all of your digital files or index them for your ImageSite document management system.

SimpleIndex automates the indexing process of large numbers of scanned images or electronic documents. SimpleIndex was designed to be easy to use yet offers extensive power and flexibility. It does this by providing a smooth workflow that takes the user automatically through the stages of scanning, processing, indexing and exporting documents. This design enables SimpleIndex to perform unique one-click scanning and unattended processing jobs.

## SimpleIndex Key Features and Functions

- Easy to use, one-click processing

- Enables automatic scanning, indexing and uploading to ImageSite document management.

- Capture full page OCR and targeted field index information thru "zonal" OCR directly from scanned documents for use in automated indexing and uploading to the ImageSite document management system.

- Zonal OCR (Optional Character Recognition)

- Full Page OCR

- Barcode Recognition

- OMR (Optical Mark Recognition)

- ISIS & TWAIN scan interface

- Locate moving OCR zones with pattern matching

- Extensive pattern matching options enables high success rate on OCR indexing functions.

- Index any file on your computer

- Command-line interface for easy application integration

- Auto-populate index field data from your database

- Audit batches for required documents

- Optional electronic imprinting and dates stamping document management system.

## Automating Document Capture

The two main methods for automating indexing are barcode recognition and OCR – Optical Character Recognition. Barcode recognition is faster and more accurate, but your documents must contain a barcode on the document or a cover page for this to work. OCR is able to read printed data directly from the page, which means most documents can be processed as-is. However there are many conditions that can affect the practicality of OCR that will be discussed in this section.

If your index data already exists in another database, SimpleIndex has two features that can make use of this data to automate processing. The Index Autofill feature lets you enter one key field that is used in a database lookup to retrieve matching values and fill in the remaining index fields automatically. SimpleIndex also has the ability to pre-set index values using the Command Line Interface and have a scanned document receive these indexes automatically.

## Using Barcode Recognition

Barcode recognition is the most efficient way to capture index data printed on documents. Some documents already have key information in barcode format on them. If your project is to scan new documents on an ongoing basis, it may be possible for you to redesign the forms to include barcodes. Having a barcode with index data on the document is the best case scenario, since all the index data is on the document at the time it is created in a format that can be read with near 100% accuracy.

If it is not possible to print barcodes on the document itself, an alternative is to have the person who creates the document print a barcode cover page and place it on the file before it is scanned.

Barcode recognition can also be useful when you have documents with a variable number of pages that will all receive the same index values. If it is not possible to generate an indexed coversheet for these at the time they are created, a generic barcode coversheet can be used to separate the scanned images into multi-page files, one for each document. A second process can then be used to index these images one file at a time instead of one page at a time, greatly increasing throughput.

## Using OCR

Zone OCR solutions traditionally require you to specify a region on the page where index information will be found. This region is recognized and the result is inserted into an index field. The problem with traditional zone OCR is that if the region is moved slightly due to variations in scanning, the result could contain extra neighboring characters or cut off desired characters. This limits the usefulness of traditional zone OCR to documents where the index value is in the exact same place every time and has plenty of white space around it.

SimpleIndex's OCR contains many advanced features to overcome the inherent limitations of zone OCR. This is done by providing template and dictionary matching for OCR fields. These features search the OCR results for a certain pattern or list of possible values and return only the matching data.

This allows you to draw your OCR zones much larger than normal, ensuring that no matter how much the data shifts around it will always be contained within that region.

It is even possible to draw your zone around the entire page and find key information that is not printed in any fixed location. For example, a doctor's office may receive lab reports from many different labs. Each report is formatted differently, but each contains the patient's name somewhere on it. Using the dictionary matching feature with a patient name list, SimpleIndex can identify the correct patient for each lab automatically.

When implementing OCR for document automation, carefully consider the data you are trying to recognize. Is the text legible? Does it appear in a fixed location? Does it conform to a unique pattern that won't be found anywhere else on the page? Is there a list available with all the possible values for this field? Answer these questions and you will know which OCR approach is best for your application.

## Index Autofill

The Autofill feature of SimpleIndex is an easy way to associate many index fields with one document without retyping data that already exists in another application. Autofill uses a database lookup to retrieve records that match a key value entered by the user. Blank index fields are then filled in automatically with the data from this lookup. The result is a document database with many different possible search fields, of which only one needed to be entered during scanning.

The key field may be typed by the user or it may be read from the document automatically using barcode recognition or OCR. The lookup is performed either when the user changes this field or when the index values are saved. If the lookup finds multiple matching records, the user will be notified and the first set of values will be used by default.

## Distributed Document Capture

Distributed Document Capture is a key feature of SimpleIndex that can provide huge savings, fast turnaround and improved quality control for enterprise scanning applications. For large enterprises with many locations that need to capture and share data, distributed document capture with SimpleIndex is an ideal solution.

To scan documents from many locations with a centralized solution, they must first be shipped to a scanning facility where dedicated personnel are required to process the large volume of paper. SimpleIndex makes it possible to scan these documents at their original location and send them electronically to a central database where they are made available throughout the enterprise.

The savings in shipping costs alone can often justify the investment in distributed capture with SimpleIndex. Other advantages include faster turnaround, keeping the originals on-site, eliminating the need for dedicated scanning departments or costly outsourcing, and the ability to ensure the presence of required documents at the time they are collected using SimpleIndex's Batch Auditing feature.

## Centralized Administration of Job Configuration

SimpleIndex jobs are highly automated through the use of pre-defined configuration files for common scanning, sorting and indexing tasks. Once defined, job configurations provide easy to use standard operational procedures for all end users.

SimpleIndex job configurations may be hosted on a website and launched via web links. This allows administrators to have a single version of all configuration files that can be updated without having to distribute local copies to many locations. A web portal can be created with links to all of the jobs relevant to the user, as well as processing instructions and other relevant information.

## SimpleIndex Additional Features and Functions

**SimpleIndex is primarily driven by requests from the ImageSite/Engine-Box browser application. The specific functions include:**

- **One-Click Processing** – SimpleIndex is designed to streamline document scanning and indexing through a single user environment. It does this by performing all of the steps in the scanning process in sequence. SimpleIndex offers the automation you need to perform zonal OCR, full page OCR, barcode recognition, database lookups, indexing, pdf conversion and export to your ImageSite Document Management system with a single mouse click.

- **Template and Dictionary Matching** – This simplified version of what is commonly known as Unstructured Forms Processing allows you to automatically extract index information no matter where it appears on the document. Most "Zonal OCR" solutions only work if the information appears in exactly the same place on every page. SimpleIndex allows you to automate indexing as long as the field matches a unique pattern or pre-defined list of values.

- **Command Line Interface** – The Command Line Interface allows SimpleIndex to be integrated with custom software solutions with minimal programming. This interface allows you to specify all job settings, set index values and process images with a single command. SimpleIndex also has the ability to run a command line application automatically before or after each batch. These can perform file format conversions on the import and export files, upload files to an FTP server, email files or launch custom application integration.

- **Document Presence Auditing** – Some applications have a set of standard documents that are required to be present in each batch. SimpleIndex has the ability to specify a list of required documents then audit each batch at the end to ensure all are present. Any missing pages may be scanned and indexed automatically from this screen as well. This feature is designed for front desk scanning applications, where several required documents are collected and scanned when they are presented by a client.

**OMR (Optical Mark Recognition)** – SimpleIndex offers a unique new approach to determining where the first page of a new document starts. Traditionally, barcode separator sheets are inserted during document prep to mark the start of a new document. It is wasteful and time-consuming to insert them between each file, especially if the files are small. SimpleIndex takes advantage of OMR technology to provide an easier solution to this problem. Simply take a felt pen and make a black mark on the upper-left corner of the first page of each new document. SimpleIndex will scan automatically to numbered multi-page files, with a new file created each time a mark is detected. These files can then be indexed and exported with a second SimpleIndex job.