

Optical Character Recognition (OCR)

What It Is

OnBase OCR (Optical Character Recognition) is a highly accurate and easy to use document recognition package integrated seamlessly with the OnBase Document Imaging module. OnBase OCR recognizes and translates printed alphanumeric characters resident on a scanned image document, converting the image into a machine-readable text document to facilitate text searching and/or full text indexing. Creating text renditions of your image documents gives you the capability to use text searches to find a particular document, and to find specific content within a document. Users are then able to instantly locate specific words or phrases within a document or group of documents, dramatically streamlining even the most cumbersome of research tasks. Documents commonly scanned and then OCR'd include business correspondence, legal memoranda, contracts, specification sheets, patient EOBs and manuals.

How It Works

OnBase OCR utilizes the ScanSoft OmniPage® OCR engine, widely praised for its high degree of accuracy and flexibility. OnBase OCR is a full-page OCR package, with multiple output options. Text renditions can be stored in ASCII text, PDF, MSWord, or HTML file formats. Formatting can be kept, keeping the original look of the image documents, or removed entirely, keeping only the document text. You can also vary the speed of the OCR process, to perform for highest character recognition or for speed, depending on your needs and the quality of your original documents. OnBase OCR offers the flexibility of having different OCR settings saved for each document type you process. OCR can be performed on bi-tonal, grayscale and even color images, and currently supports character recognition for 18 languages.

Who Will Benefit

The OnBase OCR module will benefit any company that is scanning in paper documents, and has a need to search for specific text strings within those documents, in order to quickly locate a specific document or page within that document.

Technical Features

- Run on demand or as a scheduled job.
- Flexible OCR formatting, allows speed and accuracy to be set as needed by document type.
- Output file format selection, including ASCII text, PDF, MSWord, or HTML.
- Keep or remove formatting of OCR'd documents.
- Supports 18 languages, on the same workstation.
- Utilizes the ScanSoft OmniPage® OCR engine..

Business Benefits

- Quick and simple retrieval of documents based on content within those documents.
- Eliminates the need to index documents with every value that might be used to retrieve them.
- Go directly to the page within a document that contains the information you are looking for.
- Allows search capability within image documents, and makes those documents eligible for full-text indexing.

System Requirements

- Windows 98, NT, 2000 or XP
- SQL Server 7 or greater
- Oracle 8: v 8.0.5 or greater
- Oracle 8i: v 8.1.7.2 or greater
- Oracle 9, 9i
- Sybase SQL Server v11.5
- Sybase SQL Anywhere 5.5.04, 7.0.0 to 7.0.3 and 8.0 (No 6.0.X releases)