

OpenText Capture Document Reader

A powerful document analysis tool that offers auto-classification and metadata extraction, utilizing rule-based and self-learning technology



Automates document classification and metadata extraction



Combines rule-based and **self-learning technology** for optimal results



Provides a configurable, modular structure with load balancing features for initiating and **managing powerful capture workflows**

Document analysis demands vary widely based on the unique needs of an organization. Simple analysis can consist of associating individual documents with a business process, while a more sophisticated capture workflow often involves extracting content from multi-page, unstructured business documents. Whatever the requirement, today's organizations need a capture solution that classifies incoming documents and then extracts relevant metadata for use in business workflows.

OpenText™ Capture Document Reader meets an organization's document analysis requirements through a combination of rule-based and self-learning technology. The solution features three recognition modules and voting system capabilities that work together to achieve accuracy. Its modular, configurable architecture includes load-balancing to ensure maximum performance and scalability.

Automates document classification and metadata extraction

Document Reader structures a sequential stream of scanned images as documents, sub-documents and attachments to create a complete digital file structure. It includes a large selection of modules to perform a broad range of document analysis tasks, as well as programming interfaces to assist in integrating business logic into the recognition process to increase automation rate. Rule-based and self-learning technology also enable high accuracy rates upon initial configuration, improving accuracy over time.

Combines rule-based and self-learning technology for optimal results

The solution's three recognition modules ensure accurate results:

- OpenText™ Capture Document Extraction provides rule-based document classification and extraction.
- Adaptive classification technology (ACT) is a self-learning process for document classification and requires a sample set from each class for training.
- Adaptive read technology (ART) is a modern, adaptive document extraction process, which analyzes data entered at manual data entry workstations and can identify similar documents.

Dynamic mapping can also be utilized to locate data, even if it is in a different position on a new document. The solution's voting capabilities increase accuracy when Document Extraction and adaptive read technology are utilized together. Users start with a high automation rate based on the rules in Document Extraction, with adaptive read technology improving this rate over time.

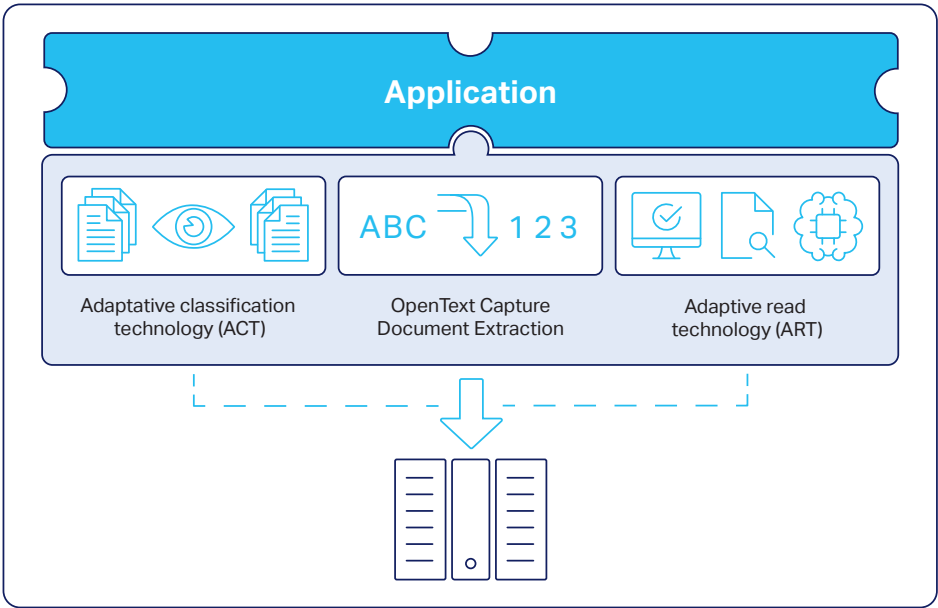
Provides a configurable, modular structure with load balancing features for initiating and managing capture workflows

Document Reader includes an integrated development environment with a graphical user interface (GUI) for configuring processing sequences in which individual modules can be parameterized and tested. A runtime system calls the individual modules, distributes the load over multiple servers within a cluster and allows an administrator to monitor and interact with the process. The solution's flexible architecture allows project-specific processing steps to be inserted at any point, and the integrated development environment supports custom processing modules. For integration with a third-party system, such as a capture workflow or mySAP™ ERP, Document Reader is presented as a subsystem with a simple XML-based interface.

A comprehensive tool for document analysis

Document Reader features a configurable framework and a wide array of recognition modules to address today's document analysis requirements. From auto-separation of multi-page documents to extraction of data from variably structured forms, Document Reader utilizes a combination of rule-based and self-learning technology to not only return high accuracy rates upon implementation, but also improve rates over time. This is accomplished in an environment that is easy to configure, manage and integrate with third-party systems.

- [Keep up to date](#)
- [Learn more](#)



OpenText Capture Document Reader utilizes a combination of rule-based and self-learning modules with the workload distributed over a server cluster. It integrates with third-party applications through an XML interface.

Feature	Description
Rule-based and self-learning technologies	Combines to offer automated document separation, classification and extraction
Voting system capabilities	Increases accuracy, allowing the system to become smarter over time
Barcode reading	Identifies and classifies content being ingested
Flexible, modular architecture	Configures through a graphic user interface (GUI)
Programming interfaces	Assists in integrating business logic
Customer processing	Supports custom processing in addition to standard modules
Processing load	Distributes over multiple servers in a cluster
XML interface	Offers easy third-party system integration, with Document Reader running as a subsystem
Asian language options	Supports the processing of Chinese (traditional), Chinese (simplified), Hong Kong Chinese, Japanese, Korean and Thai languages