



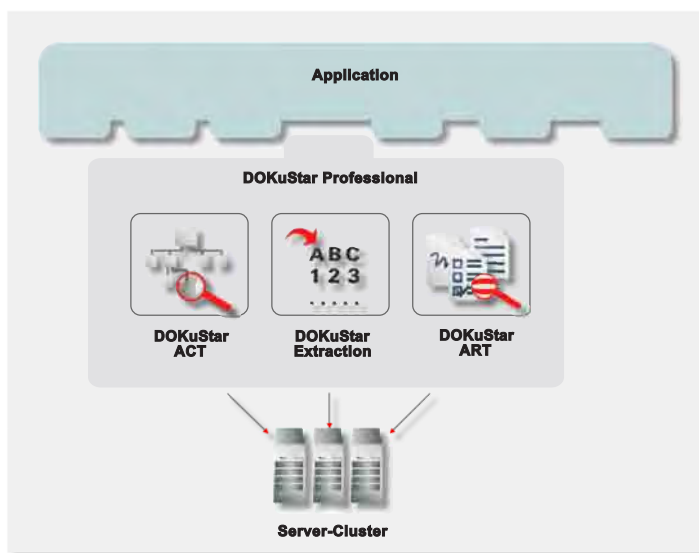
DOKuStar Professional



DOKuStar Professional – Multiple Engines for Document Analysis

Document analysis means classifying documents and extracting their user data. Automating this task increases the productivity of office work, while enhancing the transparency and reliability of business processes. DOKuStar Professional is the ultimate software for this task.

Like the proven DOKuStar Extraction, DOKuStar Professional is a recognition engine designed to be integrated as an application component. DOKuStar Professional contains DOKuStar Extraction as one of its modules, while expanding its range of functions in several dimensions.



Learning and rule-based procedures combined

DOKuStar Professional supplements DOKuStar Extraction's rule-based document analysis with the learning modules DOKuStar ACT and DOKuStar ART.

ACT stands for Adaptive Classification Technology and is a self-learning classification technology. ART is short for Adaptive Recognition Technology and is an example-based extraction technology.

Rule-based and learning procedures complement each other perfectly. DOKuStar Extraction and a suitable set of rules provide a high rate of automation when the system is being deployed, while DOKuStar ACT and ART ensure that performance continually improves. Voting based on the two sets of results, each obtained in a very different manner, leads to a further boost in recognition accuracy.

DOKuStar ACT and ART are also suitable to use from the very start: you rely completely on the automated process. Either way, DOKuStar Professional delivers unparalleled recognition rates.



Versatility: from pages to documents to file structures

The demands placed on a document analysis system can vary widely. For example, a simple task would be to associate a batch of documents with individual business processes. DOKuStar Professional's powerful, rules-based document classification lets you assign each document to a specific class. A more sophisticated task would be to extract content from multi-page, freely formatted business documents.

DOKuStar Professional relies on the proven free-form DOKuStar Extraction recognition system for this purpose. The ultimate challenge in document analysis is to structure a sequential stream of scanned images as documents, subdocuments and attachments to create a complete digital file structure. DOKuStar Professional handles this by coordinating the deployment of several modules for separation, classification, data extraction and composition, as described below.

Scripting: made to measure

The intelligent analysis of a document cannot be separated from the logic of the associated business processes. To automate, for instance, order processing, it's useful to have historical order information to assist in mapping item designations used by the customer to the company's own designations. DOKuStar Professional features a broad range of programming interfaces to assist in integrating business logic in the recognition process and increase the automation rate. These are embedded in a flexible architecture, allowing project-specific processing steps to be inserted at any point. And that's not all: DOKuStar Professional is an open platform. Thanks to modern plug-in technology the integrated development environment recognizes and supports custom processing modules in addition to the included modules.

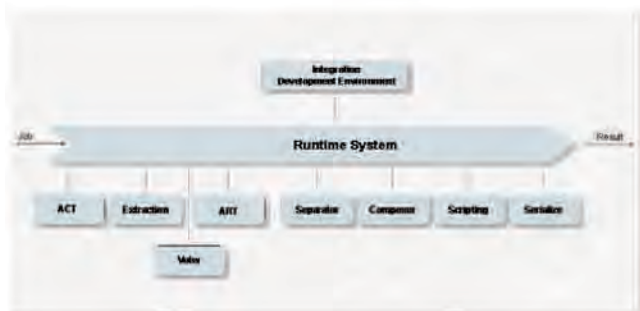
Load balancing: limitless CPU power

Despite the flexibility, versatility and power of DOKuStar Professional, the integration of the software is simple. DOKuStar Professional presents itself to the leading system – a capture workflow or a business application such as mySAP ERP – as a subsystem with a simple XML-based interface.

The full complexity of the application, including load balancing and error recovery, is encapsulated behind this interface. With the load-balancing option enabled, DOKuStar Professional distributes jobs across server clusters and monitors them. The only thing that the leading application notices of all of this is that it receives highest-quality document analysis results very promptly.

Flexibility: a hallmark of modern architecture

DOKuStar Professional resolves the conflict between ease of use and highest possible flexibility with its modern .Net®-based architecture. The Framework supports the free configuration of individual modules and expansion with additional modules. The diagram below illustrates the architecture.



The **modules** perform the actual work in processing the documents. DOKuStar Professional comes with an impressive selection of modules for the broad range of document analysis tasks described below. The Framework provides a simple means to add additional modules or implement application-specific functions via the appropriate programming interfaces.

The **runtime system** is responsible for three tasks. It calls the individual modules, distributes the load over multiple servers within a server cluster, and lets the administrator monitor and interact with the process. All of its functions must meet high standards to support the wide range of possible applications. A variety of processing sequences can be defined for an application, for example for service providers who need to vary the processing of their documents according to clients. The sequence variations are also evenly distributed in the server cluster. The load balancing goes a step further, however – the server cluster can also be shared by other applications to



ensure optimal utilization of hardware resources when processing high volumes and using a variety of applications.

The **integrated development environment** is a tool for application developers. It features a graphical user interface in which processing sequences can be configured and the individual modules parameterized and tested. Modules developed by the user can also be graphically represented on the graphical user interface thanks to plug-in support. The entire application is set up for maximum performance in fast work cycles. Finally, a deployment mechanism automates the implementation of the finished application in the production system.

The **job interface** represents the custom application vis-à-vis the leading system. During productive operation, DOKuStar Professional acts as a self-contained subsystem that is seen as a single unit by the leading system. All of the complexity arising from a variety of applications and distribution on multiple servers is hidden behind a single simple interface. The XML-encoded jobs can be processed synchronously or asynchronously, depending on whether the leading system is waiting for the results or needs to continue working in the meantime.

The modules: processing engines

DOKuStar Professional defines an application as a number of processing chains (pipelines). Each pipeline is a sequence of modules, in which individual modules may also appear repeatedly. Data is exchanged between individual modules as so-called runtime documents.

DOKuStar ACT is a self-learning process for document classification. A sample sentence from each document class is sufficient to train the system. The software learns what the characteristic features of this sample are, and later automatically recognizes other documents of the same class. The process is extremely easy to use during the training phase and is quick and reliable in operation.

DOKuStar Extraction is the established standard product for rules-based document analysis. It classifies individual pages according to classes and types and extracts application data using Captaris Document Technologies' proven free-form recognition. In an application, DOKuStar Extraction guarantees high recognition rates already at the beginning of pro-

ductive deployment. DOKuStar Professional sends individual jobs, which can vary from page to page, to DOKuStar Extraction for every page of the document being processed.

DOKuStar ART realizes a modern, adaptive document extraction process. ART stands for Adaptive Read Technology. Data acquired from documents by users is analyzed and stored in the Repository. DOKuStar ART can then identify similar documents. Dynamic mapping ensures that the required data is found in the current document, even if the data is in a different position or a changed environment.

The **Voter** comes into play whenever DOKuStar Extraction and DOKuStar ART are used simultaneously in DOKuStar Professional. The parameterization of DOKuStar Professional controls how the results of the two recognition processes are combined, with rules applicable on a field-specific basis.

The **Separator** divides the homogenous page stream into individual documents. The recognition results of DOKuStar Extraction are applied to the separation. Not only may the detected document type, but also individual data items extracted from the document be used for this purpose.

The **Composer** then groups the recognition results of individual pages as the results for the document. Similar data may be found in multiple instances over several pages, requiring a decision on the result that will apply for the document as a whole. In other cases, data may be distributed over several pages of a document, as is often the case with tables. These situations require a merger of the individual data records.

The **scripting modul** is the ideal place for application-specific programming. The object model established and proven in DOKuStar Validation, in which the data of the runtime document is converted, serves as the programming interface. Multiple instances of the scripting module can be inserted in a variety of places in the process.

The **XSL and Serialize modules** form the end of the processing chain. With the XSL module, data can easily be converted at the interfaces between modules using XML technology, be it to generate a certain result format or to eliminate temporary data (intermediate results). The Serialize module stores the results as a file in the file system in case the leading system does not load the results to memory.



Good things come in threes: product versions

DOKuStar ACT is included as standard in DOKuStar Professional. Either DOKuStar Extraction or DOKuStar ART can be used, or both together. Thus, three configurations are available:

- DOKuStar Professional with Extraction is a logical enhancement of the basic DOKuStar Extraction product featuring multi-page support, load distribution and scripting options.
- DOKuStar Professional with ART is an innovative entry into document content acquisition. Rather than defining extraction rules in advance, ART lets the system optimize itself during productive use.
- DOKuStar Professional with Extraction and ART combines the best of both worlds. Users can start with a high automation rate based on DOKuStar Extraction and improve it further with adaptive learning thanks to DOKuStar ART.

© 2008 All rights reserved. No part of this publication may be reproduced, transmitted, transcribed, stored in a retrieval system, or translated into any language in any form by any means without the written permission of Captaris. All brand names and trademarks are the property of their respective owners.

Order Nr. L93061-N159-X-3-7618, 415 209, 02/08, printed in Germany.



Headquarters (Germany)

Captaris Document Technologies GmbH
Max-Stromeyer-Str. 116
78467 Konstanz

Phone (++49) 7531 87-4500
Fax (++49) 7531 87-4567
www.Captaris-dt.com

US Office

Captaris, Inc.
Captaris Document Technologies
4340 East West Highway, Suite 201
Bethesda, MD 20814

Phone (301) 652 9732
Fax (301) 652 7088
www.Captaris-dt.com