

ABBYY FlexiCapture Engine 12

The Most Comprehensive SDK for Data Capture

What is FlexiCapture Engine?

ABBYY FlexiCapture Engine is a comprehensive software development kit for developers who need functionalities for automated data capture, form processing or intelligent document classification in their applications. After integrating this SDK via an API, information on scanned or photographed documents can be automatically extracted and used. These documents can be standardised forms, such as questionnaires, as well as documents with various layouts, such as delivery notes, invoices or contracts. In addition, large batches of pages can be automatically merged to original documents and sorted into classes to speed up their processing. With the new functionalities, the value of own applications can significantly increase.

Extend Existing Applications or Build Automated Data Capture Solutions

ABBYY FlexiCapture Engine delivers a full-range of state-of-the-art technologies for data extraction, document classification and indexing. Ready-to-use components for exporting data into backend processing and archiving systems can be easily integrated into a single application.

Intelligent Document Processing

Automated data extraction from fixed forms, processing of semi-structured and unstructured documents and document classification in one SDK.

Advanced Processing of Multi-Page Documents

Unique technology to process documents with multiple pages easily, even documents containing tables that continue on the next page.

Reliable Platform for Processing

Powerful SDK, compliant to industry data capture standards for data processing in enterprise projects.

Automated Data Extraction into Backend Applications

Using FlexiCapture Engine, extracted data can be exported into any backend application via standard interfaces.

Processing of Multilingual Documents

Support for multiple languages to allow development of localised applications in the international environment.

Auto-Learning Features

The Auto-Learning API and a source code sample offer a fast way to train in data extraction from simple document types via a point-and-click user interface. Once the training is complete, the data will be automatically extracted from similar documents.

Cost Savings

Reduce costs of data extraction by deploying automated data capture capabilities.

Efficient Development

A set of standard components enables fast integration, thus shortening development cycles.

High Quality Data Output

Highly accurate recognition technologies combined with rule-based automated verification methods ensure optimised and validated data output.

Tight Integration with Applications

FlexiCapture Engine enables developers to ensure that data capture technologies are perfectly integrated into their applications.

KEY FUNCTIONALITY

- Integration of intelligent data capture to process forms and questionnaires
- Document classification and automated document separation
- Powerful data extraction for semi-structured and unstructured documents
- Verification and monitoring methods for optimised data output
- High-quality recognition technologies for OCR, ICR, OMR, 1D and 2D Barcodes
- Language support for 205 OCR and 126 ICR languages
- Automated data export to backend applications
- PDF, MRC PDF & PDF/A document export for archiving

BENEFITS FOR DEVELOPERS

- Enhance your applications with data capture components
- A unique and powerful set of features and API for data extraction
- Ready-to-use SDK components to speed up development processes
- Full control over document processing, extracted data and document routing
- Compatibility with FineReader Engine 12 API, which can be licensed within FlexiCapture Engine for integration of document conversion

Development Overview

How to Implement Data Extraction Functionalities

The integration of data capture functionalities with FlexiCapture Engine is a two-stage process. Stage 1: First, the SDK needs to be integrated into your application. For this, developers can use the SDK components provided. Stage 2: For the subsequent processing of different document types within a project, the relevant data extraction fields need to be defined. For standardised forms these fields are defined in document templates. For documents that have varying layouts, such fields must be described using special tools in order to locate them on the document. Professional design tools are available for those tasks: Document Template Editor to create templates for standardised structured forms and FlexiLayout Studio to describe data capture fields on unstructured or semi-structured documents. The provided API helps to train in the data extraction process on one-page documents with own GUI.

Integration in Custom Applications and Processing

FlexiCapture Engine has an easy-to-use API. Development can proceed in C#, C++, Visual Basic .Net, Delphi, or any other tools supporting COM components. The following steps show how your application can be extended with Data Capture functionalities:

Step 1 - Integration

Integrate FlexiCapture Engine DLLs into your application, workflow or business system.

Step 2 - Project Selection:

Load a predefined set of document templates that have been created via the Auto-Learning API or developed and tested with FlexiLayout Studio.

Step 3 - Document Processing:

Load the documents and PDFs you wish to process. Document separation, classification and recognition are executed automatically according to the pre-defined document templates.

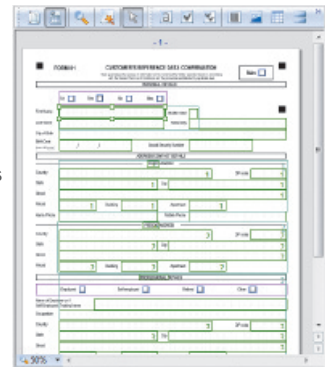
Step 4 - Data Verification:

The Data Verification API of FlexiCapture Engine provides access to the internal recognition results. The Visual Components, provided as a source code sample, allow fast and easy implementation of a GUI for verification.

Step 5 - Data and Document Handling:

Extracted data can be processed in your application, sent to other applications or exported to XML, XLS or CSV. The separated documents can be exported as image files, as well as searchable PDFs or PDF/A files.

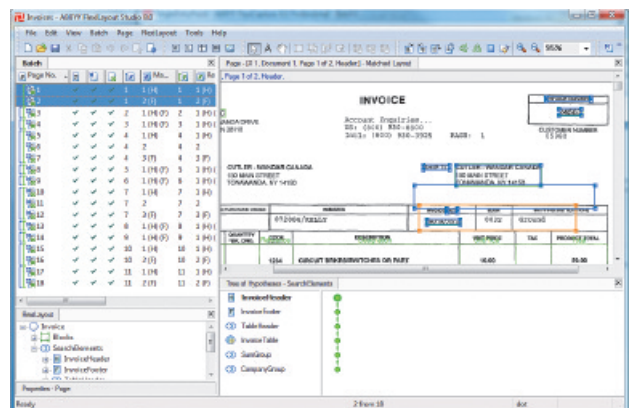
- 3) Define the different recognition areas in the Template Editor, e.g. text blocks, tables, checkmarks, checkmark groups, barcodes and pictures.
- 4) Set up the recognition properties for each area, e.g. OCR, ICR and attach data type definitions, dictionaries and verification rules.



Development of FlexiLayout Templates for Documents with Varying Layouts

The FlexiLayout Studio allows the creation of rules for detecting data fields on documents that cannot be matched to standard templates because they may have various layouts.

- 1) Load a selection of documents with different layouts into FlexiLayout Studio.
- 2) Define specific elements that allow document to be identified and that can be used for orientation within the document, e.g. text strings, lines or spaces between elements.
- 3) Define search elements for the data you are looking for e.g. text, numbers, date, tables, length of the string, the set of characters, one or multiple words or one or multiple lines. Additionally, these elements are set into a relation with elements set up in the previous step. For example: 'Find text below "Invoice Number" that is between 8 and 16 characters long.'



Development of Document Templates for Forms

The Document Template Editor allows fast and intuitive development of document templates to process static, fixed forms.

- 1) Load all pages and attachments of the form into the Template Editor.
- 2) Define the elements that will be used to match documents to this template: anchors, static texts or separators.

Functionality Overview

ABBYY FlexiCapture Technology for Data Capture

Unique Set of OCR, ICR, OMR and Barcode Recognition Technologies

FlexiCapture Engine's core recognition technologies process 205 languages in OCR, 126 languages in ICR as well as checkmarks (OMR) and barcodes.

Processing Different Document Types in One Stream

FlexiCapture Engine contains ready-to-use components which enable developers to integrate automatic extraction of data from a stream of forms and documents. Automatic document type detection allows processing of different documents (structured, semi-structured, unstructured) via document templates and FlexiLayouts.

Document Separation

FlexiCapture Engine offers two types of document separation. Simple separation is provided via blank pages and barcodes. Advanced document separation is possible via intelligent document definitions. Even template matching can be performed without any manual work.

Document Classification

Once a document matches a document template or a FlexiLayout, the relevant document type is automatically identified, e.g. customer contract, invoice or letter.

Form Processing

FlexiCapture Engine is perfectly designed to process any kind of fixed forms, like questionnaires, surveys and/or coupons.

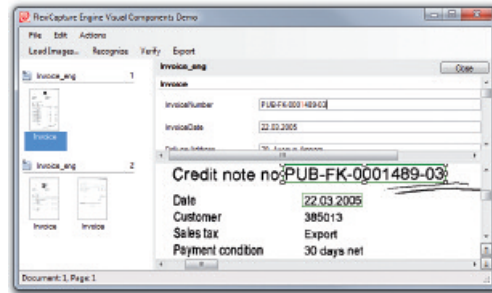
Advanced Processing of Multi-Page Documents with Different Layouts

Most business documents, such as price lists, contracts or surveys consist of several pages. With FlexiCapture Engine, developers get a unique technology to process multiple page documents easily, even documents with tables across several pages.

Data Verification

FlexiCapture Engine offers effective and convenient instruments for checking the validity of captured data and detecting recognition errors:

- Automatically: If necessary, data can be automatically checked. This includes a simple area check and a comparison of the results with values in a database. Moreover, scripting can be used to create own sophisticated rules.
- By a user: The API of FlexiCapture Engine allows the development of own user interfaces. With the Visual Components (Source Code Samples) developers can offer users an effective, ergonomic user interface.



Data & Document Export

ABBYY FlexiCapture Engine is able to index a document with relevant data for export into backend applications and converts analysed documents into searchable PDFs, highly compressed (MRC) PDFs or PDF/A files for archiving.

Java Wrapper

FlexiCapture Engine provides ready-to-use Java classes for the Engine library. These Java classes can be used directly from Java applications.

Auto-Classification

FlexiLayout Studio features Auto-Classifier, a function that combines raster and content classification. The technology is easy to set up and enables fast processing. Classification training is also available via API.

Document Template Development

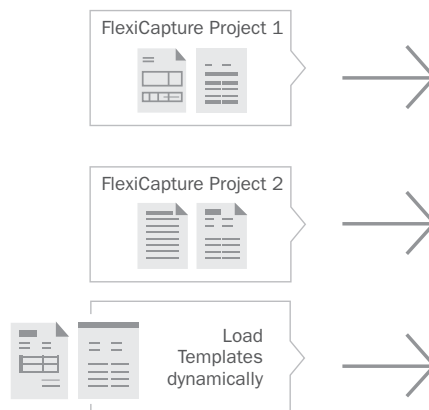
Fixed Form Templates:
Questionnaires, Surveys, Forms, ...



Flexi Layouts:
Invoices, Contracts, Letters, ...



Document Collection



Application with FlexiCapture Engine



Specifications and Licensing

SPECIFICATIONS

System Requirements

- PC with x86 processor (1 GHz or higher)
- Operating system:
32-Bit: Windows 10, Windows 8, 8.1, Windows 7,
64-Bit: Microsoft Windows Server 2012, 2012 R2, 2008 R2, Windows 10, Windows 8, 8.1, Windows 7
- Memory: min. 1 GB RAM

Cloud Platforms

- Microsoft Azure
- Amazon EC2

Input Formats

BMP, PCX, DCX, JPEG, JPEG 2000, JBIG2, PNG, GIF, TIFF, DjVu, scanned and digitally created PDFs

Output Formats

Editable formats

- XLS, DBF, CSV, TXT, XML

Searchable formats

- PDF and PDF/A with MRC compression, tagged PDF

Image formats

- BMP, PCX, DCX, JPEG 2000, JPEG, TIFF, PNG, PDF, PDF/A

Development

FlexiCapture Engine API supports the COM standard and can be easily used in C#, C++ (raw and COM), Visual Basic .Net, Delphi or others. The distribution package contains code samples for C#, C++ (raw and COM), VB.Net, Java, Delphi and scripting languages like VBS, JS and Perl.

Further information online:

<http://www.abby.com/data-capture-sdk/technical-specifications/>

ABBYY Licensing Policy

ABBYY FlexiCapture Engine is sold via a flexible, modular licensing policy which allows developers to select the best combination of tools and pricing options for their project. Licensing is offered as:

Development Licence

Allows the integration of ABBYY FlexiCapture Engine into own applications and functionality testing. The licence also includes the FlexiCapture Standalone application with the Flexilayout Studio tool. These tools allow the design of templates for processing standardised forms and documents with various layouts.

Runtime Licences

Allow the distribution of applications that include FlexiCapture Engine functionalities. Runtime licences differ in their functionality, in the number of pages that can be processed, or in the number of CPU cores that can be used for page processing (without a page counter). By default, ABBYY Europe ships network licences. Instances, running in the LAN on virtual or physical computers, can use this network licence via a central Network Licence Manager.

Add-on Modules for Runtime Licences

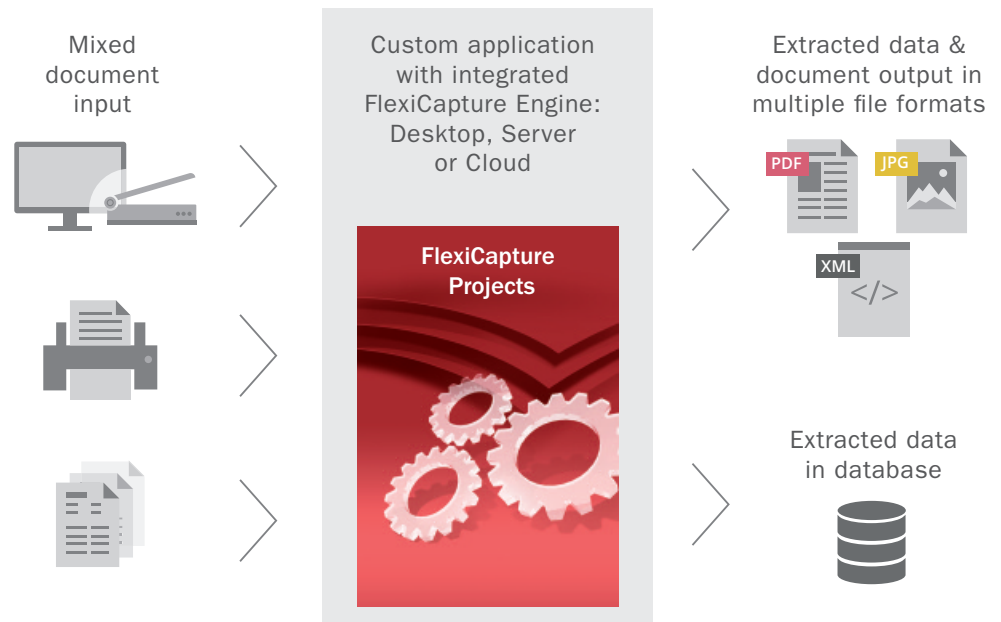
Runtime licences can be expanded by adding specific functionalities, for example additional OCR languages or the PDF export function. In addition, the ABBYY FineReader Engine API for full-text OCR and document conversion can be licensed with FlexiCapture Engine.*

Software Maintenance, Training and Professional Services

To secure the success of your products, ABBYY offers additional support, training and education programs for all its products. If you would like professional support or consulting, please contact us.

* FineReader Engine API is included in the licence model of ABBYY Europe; this does not apply to other sales areas.

ABBYY® FlexiCapture Engine Document Processing



ABBYY®

Headquarter Europe
ABBYY Europe GmbH
Landsberger Str. 300, 80687 Munich, Germany
Tel: +49 89 69 33 33 0
sales_eu@abby.com
www.ABBYY.com