ABBYY®

ABBYY® FlexiCapture® Engine 12

The tool you want for the data they need

What is FlexiCapture Engine?

ABBYY FlexiCapture Engine represents the latest generation of ABBYY's data capture SDK technology. It enables developers to quickly develop solutions for extracting data from fixed and semi-structured forms and documents. FlexiCapture Engine is ideal for system integrators and service providers who develop integrated data capture solutions.

Extend Existing Applications or Build Automated Data Capture Solutions

ABBYY FlexiCapture Engine delivers technologies for data extraction and indexing. Ready-to-use components for exporting data into backend processing and archiving systems can be easily integrated into a single application.

Smart Document Processing
Automated data extraction from fixed forms and processing of semi-structured documents in one SDK.

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

Processing of Multilingual Documents
Support for multiple languages to allow development of localized applications in an international environment.

Automated Data Extraction into Backend Applications
Using FlexiCapture Engine, extracted data can be exported into any backend application via standard interfaces.

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

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.

High Quality Data Output
Highly accurate recognition technologies combined with rule-based automated verification methods ensure optimized and validated data output.
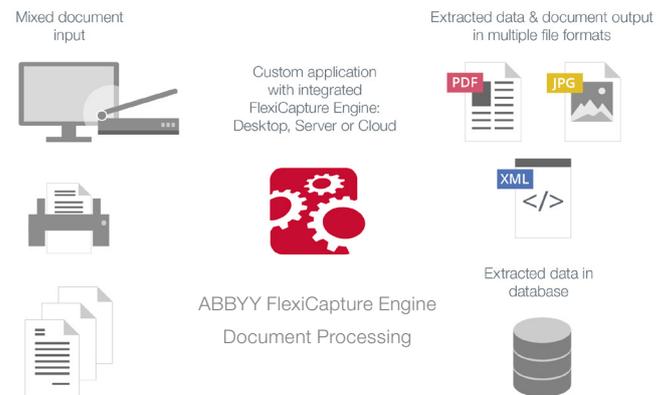
Development Overview
How to Implement Data Extraction Functionalities

How to Implement Data Extraction Functionalities

The integration of data capture functionalities with FlexiCapture Engine is a two-step process:

- first, the SDK needs to be integrated into your application,

- second — for the subsequent processing of different document types within a project, the relevant data extraction fields need to be defined.

For standardized 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 standardized 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 its own GUI.

Mixed document input

Custom application with integrated FlexiCapture Engine: Desktop, Server or Cloud

Extracted data & document output in multiple file formats

ABBYY FlexiCapture Engine Document Processing

Extracted data in database

Integration in Custom Applications and Processing

FlexiCapture Engine has an easy-to-use API. Development can proceed in C#, C++, visual Basic .net, 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 API for creation document definitions for fixed forms and semi-structured documents with labeled fields 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.

Step 5   Data and Document Handling:

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

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.

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 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 documents 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.

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 207 languages in OCR, 112 languages in ICR as well as checkmarks (OMR) and barcodes.

Processing Different Document Types in One Stream FlexiCapture Engine contains ready-to-use tools 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) via document templates and FlexiLayout.

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.

Form Processing

FlexiCapture Engine is perfectly designed to process any kind of fixed forms, like question-naires, 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 docu-ments 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 your own sophisticated rules.

- By a user: the API of FlexiCapture Engine allows the development of your own user interfaces.

Data & Document Export

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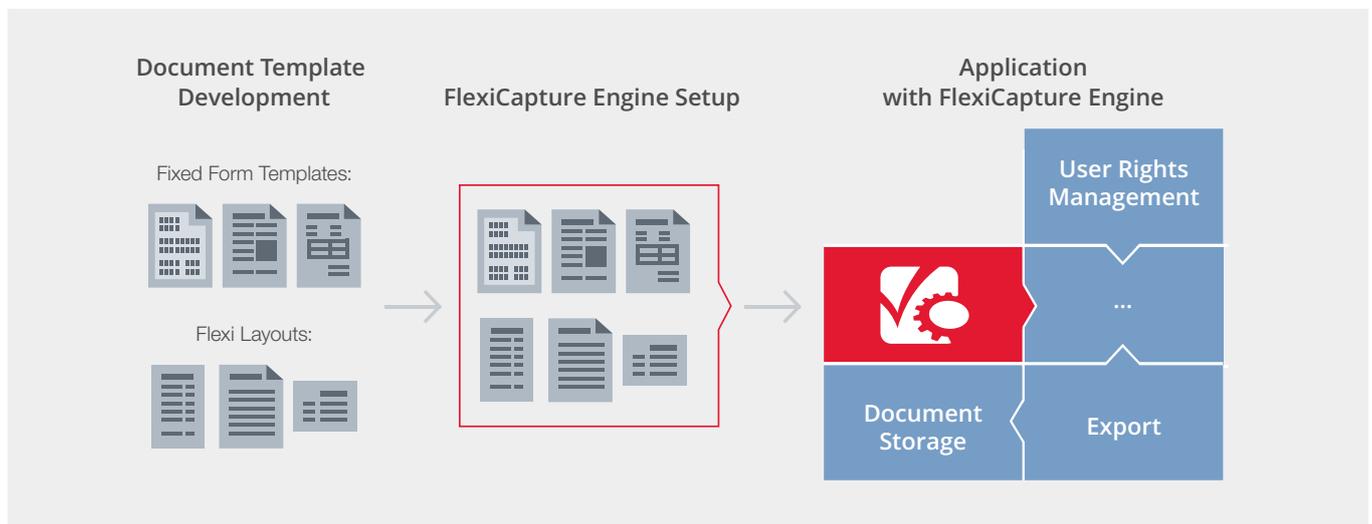
Java Wrapper

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

KEY FUNCTIONALITY

- Integration of intelligent data capture to process forms and questionnaires
- Automated document separation
- Data extraction for fixed forms and semi-structured documents
- Verification and monitoring methods for optimized data output
- High-quality recognition technologies for OCR, ICR, OMR, 1D and 2D Barcodes
- language support for 207 OCR and 112 ICR languages
- Automated data export to backend applications
- PDF export — Searchable PDF, PDF/A, MRC compression

BENEFITS FOR DEVELOPERS

- Enhance your applications with data capture components
- Full control over document processing, extracted data and document routing
- Compatibility with FineReader Engine 12 API.

Specifications and Licensing

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 License

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

Runtime Licenses

Allow the distribution of applications that include FlexiCapture Engine functionalities.

Runtime licenses differ in their functionality, in the number of pages that can be processed.

Add-on Modules for Runtime Licenses

Runtime licenses 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 is 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.

If you have additional questions, contact your local ABBYY representative listed under www.abbyy.com/contacts

© 2019 ABBYY Production LLC. ABBYY FlexiCapture Engine 12 © <year of v.12 release> ABBYY Production LLC. ABBYY, FlexiCapture, FlexiLayout, FineReader are either registered trademarks or trademarks of ABBYY Software Ltd. All other product names, trademarks and registered trademarks are the property of their respective owners.

SPECIFICATIONS

System Requirements

- PC with x86 processor (1 GHz or higher)
- Operating system: Windows® server 2016, 2012, 2008 R2 Windows 10, 8.1, 8, 7
- Memory: for processing one-page documents: At least 400 MB RAM, 1 GB RAM recommended for processing multi-page documents: at least 1 GB RAM, 1.5 GB RAM recommended

Cloud Platforms

- Microsoft® Azure®:
  Azure App Service
  Azure Cloud Services
  Azure Service Fabric
  Azure Virtual Machines
- Amazon EC2®

Input Formats

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

Output Formats

Editable formats: XIS, 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++, Visual Basic, .NET, or any development tool supporting COM components. The Engine can be adapted for use in scripting languages like VBS, JS, and Perl.

Further information online: http://www.abbyy.com/data-capture-sdk/technical-specifications/

ABBYY®

WWW.ABBYY.COM