

ABBYY FlexiCapture™ 1.5

Intelligent Technology for Semi-structured Forms and Documents



Overview: An Answer for Semi-Structured Document Processing

Each day companies around the world struggle with the monumental task of processing millions of paper documents – all of which need to be input into the computer in the quickest, most cost-effective way possible. While today's form processing systems have advanced significantly, the available options for processing semi-structured documents, or "flexible forms," have remained limited.

Traditional methods for capturing semi-structured documents have required many hours, days, or months of customisation – creation of multiple form templates for each form layout, form redesign, and technology adaptation. Or alternatively, manual input by an operator. ABBYY now offers an alternative.

ABBYY FlexiCapture Studio provides an easy-to-use solution for processing documents that contain similar sets of information, but have varied form layouts and structures. Designed as a companion product for ABBYY FormReader and ABBYY FineReader Engine, ABBYY FlexiCapture Studio enables the creation of intelligent form layouts that work together with ABBYY recognition software to detect key form field data despite variations in layout and placement. It is an ideal solution for processing a wide variety of varied, semi-structured documents in the following areas:

Finance/Banking

- Bank Payment Orders
- Financial Statements
- Credit Notes
- Loan Documents
- Mortgage Documents
- Securities
- Cheque
- Statements of Account

Accounting

- Invoices
- Purchase Orders
- Payment Slips

Transportation

- Shipping Orders
- Delivery Notes
- Customs Declarations

Legal

- Template-based Contracts
- Lease Agreements

Insurance

- Benefit Application Forms
- Claim Forms
- Expertises
- Contracts

How FlexiCapture Studio Works: The FlexiLayout

FlexiCapture Studio extends the data capture and recognition capabilities of ABBYY FormReader and ABBYY FineReader Engine beyond traditional form processing. It achieves this by enabling simple creation, testing, and debugging of a FlexiLayout, a formal description, or "map" of the fields and variable layouts of a certain set of forms.

Customers can work with FlexiLayouts by:

- **Using the FlexiCapture Studio interface** – the intuitive dialogues require no coding or programming experience.
- **Fine-tuning the FlexiLayout** – using FlexiCapture's special template creation programming language.

FlexiLayouts describe the logic of data extraction. Each FlexiLayout consists of user-defined **elements** and **blocks**. **Elements** in the FlexiLayout serve as a set of rules or descriptions of objects and their relationship within the document. They help ABBYY recognition technology find the necessary data fields during form processing and document recognition. Elements include static text, character strings, white spaces, etc. **Blocks** describe the elements from which the data must be captured – they are the areas from which ABBYY recognition technology must capture data.

High accuracy and speed is achievable because FlexiCapture uses a unique algorithm to detect data fields by analysing whole sets of form elements. This approach is dramatically different from that of traditional form processing technology which analyses separate elements one by one. By analysing a set of elements as a group, FlexiCapture can increase accuracy by viewing elements "in context" and speeds up the steps needed to process the entire form.

A completed and tested FlexiLayout is easily exported into one of the ABBYY products – FormReader or FineReader Engine – for data capture of semi-structured documents, no matter how complex the layout may be.

ABBYY FlexiCapture: Intelligent Technology

How does a human operator enter data from semi-structured forms into a computer? When looking for data, the operator looks at the entire form as a whole and then analyses various form elements and individual pieces of information such as: field captions, or units of measurement. The operator then looks at information surrounding the form fields to determine which information to enter into which part of the database.

ABBYY FlexiCapture uses artificial intelligence to enable ABBYY recognition technologies (ABBYY FormReader and ABBYY FineReader Engine) to act in a similar fashion. FlexiCapture is based on the principles of Intelligent, Purposeful, and Adaptable (IPA) recognition. IPA imitates the ways in which humans recognise objects. With user-defined descriptions, FlexiCapture guides ABBYY recognition engines toward finding the required fields on forms with varied layouts. As a result, the recognition software can find a key field and piece of data, such as "date," on various different forms – no matter where it is located on the paper (at the top left, bottom right, etc.).

Key Functionality

Interface

- Visual environment for easy FlexiLayout creation, editing, and testing
- "Tree View" option for easy viewing of FlexiLayout blocks, hypotheses, and properties
- Customisable program windows
- Drag-and-drop functionality

FlexiLayout Objects

- Block (objects that contain required data) types: text, tables, barcodes, checkmarks, and pictures.
- Element (objects that help in locating blocks) types:
 - simple elements (static text, separators, white gaps, barcodes, character string, text fragment, object collection, table, date, phone, currency)
 - compound elements (simple elements joined by AND)

FlexiLayout Creation

- Pre-recognition in 172 OCR languages, with the ability to select multiple language combinations
- Intuitive dialogs for setting element properties
- Specification of element properties based on type:
 - text to be searched with possible alternatives and numbers of errors
 - columns, rows, header and footer properties for tables
- Selection of search constraints with absolute search area coordinates and relative position over other elements
- Advanced code writing settings for FlexiCapture programming language such as automatic check of error message syntax
- Ability to set one-to-many relationships between elements
- Easy FlexiLayout error checking during creation process

FlexiLayout Testing and Adjustment

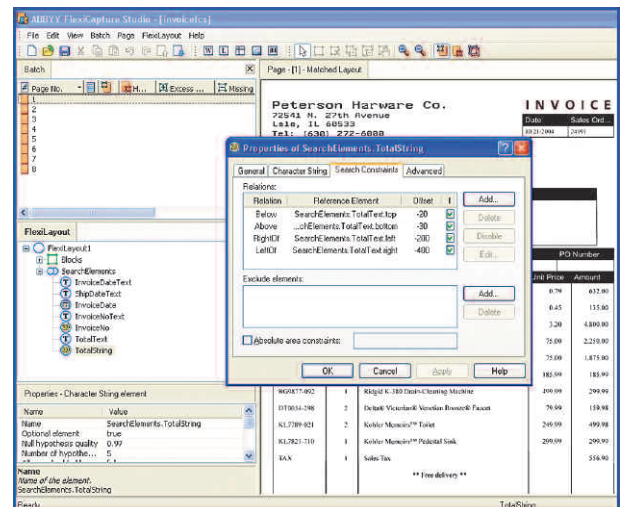
- On-the-fly testing and adjustment of FlexiLayouts using a batch of sample images
- Formulation of hypotheses (assumptions that the detected objects correspond to the elements) that are:
 - Created for whole set of elements, (as opposed to single element)
 - arranged in a tree-like structure and ranked by quality
 - analysed with best hypothesis selected and marked
- Visualization of relationship between a possible hypothesis and ^ FlexiLayout element
- Usage of null hypotheses when optional elements cannot be found
- Grouping of hypotheses in clusters
- Display of hypotheses based on element types and search results (found/not found).
- Easy testing of FlexiLayouts by comparison of results with a reference layout created by:
 - drawing elements manually
 - correcting the results of matching

Pricing and Availability

ABBYY FlexiCapture Studio 1.5 can be purchased for use with FormReader 6.5. The product is also available for use with ABBYY FineReader Engine 8.0. At least one runtime licence of the base product (FormReader or FineReader Engine) is required to process forms using FlexiCapture. For detailed pricing please contact your local sales representative.

System Requirements:

- PC with Intel Pentium®, Celeron®, Xeon™, AMD K6/Athlon™/Duron™/Sempron™ or compatible processor with a minimum clock speed of 350 MHz
- Operating System: Windows® XP, Windows® 2000 Min. Memory 128 MB RAM
- Min. 128 MB Hard disk space Video card and monitor (min. resolution 800x600)
- Keyboard, mouse or other input device
- USB Port for a Licence-Donlge



Benefits for Developers, Integrators and Service Bureaus

- Expands your business options.** Provides a flexible tool for delivering reliable and effective solutions for semi-structured forms processing.
- Supports a wide range of semi-structured documents and markets.** Can define layouts of different complexities – invoices, documents with special headers, pages with tables.
- Ease of use.** Versatile and visualised interface enables use without the need for specialised programming skills.
- Fast deployment.** Simple tools and pre-packaged samples expedite set up and testing.
- Cost-effective.** Competitive pricing and ease-of-use limits the amount of resources required to develop powerful solutions.
- Proven technology.** FlexiCapture is based on more than 4 years of development working closely with large enterprise customers on flexible form projects. Products based on ABBYY FlexiCapture technology are already in use by major financial institutions across Eastern Europe.

© 2005 ABBYY Software Ltd. All rights reserved. ABBYY, FINEREADER, and ABBYY FineReader are registered trademarks of ABBYY Software Ltd. All other trademarks are the property of their respective owners.