

ABBYY® FlexiCapture® 10



Take the Data. Leave the Paper.

ABBYY FlexiCapture 10 là giải pháp tự động bóc tách thông tin trên tài liệu giấy với độ chính xác cao. FlexiCapture tự động phân tách và phân loại tài liệu; đặc biệt, với khả năng nhận dạng nhiều ngôn ngữ (trong đó có tiếng Việt) FlexiCapture tự động trích chọn và cung cấp cho người sử dụng các dữ liệu cần thu thập một cách nhanh chóng. Giải pháp ABBYY FlexiCapture 10 dễ dàng triển khai và dễ dàng sử dụng.

Kiến trúc mềm dẻo của ABBYY FlexiCapture 10 cho phép xây dựng giải pháp để đáp ứng các nhu cầu đa dạng khác nhau - từ các hệ thống độc lập cho các công ty, văn phòng vừa và nhỏ có ngân sách hạn chế cho đến các giải pháp chạy trên máy chủ có khả năng mở rộng cao cho các dự án lớn của doanh nghiệp và chính phủ. Ngoài ra, ABBYY FlexiCapture 10 có thể tích hợp với các hệ thống quản lý thông tin và vào các quy trình nghiệp vụ chuyên biệt để nâng cao hiệu quả tổng thể và giảm chi phí.

Tự động bóc tách thông tin với độ chính xác cao

Khả năng tự động bóc tách thông tin cho phép giảm đáng kể các thao tác thủ công, qua đó rút ngắn thời gian xử lý và giảm thiểu chi phí. Công nghệ nhận dạng với độ chính xác cao cùng với các chức năng soát lỗi tự động đảm bảo cung cấp dữ liệu chính xác trong thời gian ngắn.

Tích hợp công nghệ tiên tiến nhất

Phần mềm sử dụng các thuật toán trí tuệ nhân tạo tiên tiến khi phân tích các tài liệu để tự động phân loại và bóc tách thông tin.

Xử lý khối lượng lớn

Hỗ trợ xử lý với nhiều CPU, có kiến trúc khách - chủ và cơ chế cân bằng tải đảm bảo dễ dàng mở rộng cho các nhu cầu cao hơn và phân tải tốt hơn trên mạng.

Khả năng chịu lỗi

Hỗ trợ Microsoft Cluster, đảm bảo hoạt động của hệ thống được ổn định và ngăn ngừa mất dữ liệu trong trường hợp có lỗi hệ thống.

Các ứng dụng soát lỗi đa dạng

ABBYY FlexiCapture cho phép phân bổ công việc một cách hiệu quả và giảm chi phí nhờ sử dụng ứng dụng web trong soát lỗi và nhập dữ liệu.

Dễ dàng sử dụng

Chức năng học tự động làm đơn giản hóa quá trình phân loại tài liệu và thiết kế biểu mẫu.

Công cụ tích hợp và tùy biến linh hoạt

Giải pháp bóc tách dữ liệu có khả năng tùy biến và tích hợp chặt chẽ với các quy trình nghiệp vụ hiện có của khách hàng bằng cách sử dụng các đoạn mã script hoặc Web service API.

Chi phí thấp, thu hồi vốn nhanh

Hệ thống dễ dàng triển khai, dễ dàng sử dụng, loại bỏ các thao tác thủ công trong các quy trình nghiệp vụ giúp khách hàng nhanh chóng thu hồi vốn đầu tư.

Hỗ trợ kịp thời

Là một công ty quốc tế, ABBYY có văn phòng đại diện và đối tác trên toàn thế giới. Do đó, các khách hàng của ABBYY được đảm bảo sẽ nhận được sự hỗ trợ kỹ thuật:

- Kịp thời
- Bằng ngôn ngữ bản địa
- Trong bất cứ tình huống nào

Mọi lĩnh vực

Mọi loại tài liệu

Mọi loại dữ liệu

Giải pháp tự động phân tách, phân loại tài liệu và bóc tách dữ liệu được ứng dụng trong nhiều lĩnh vực, xử lý nhiều loại tài liệu:

Y tế

Đơn thuốc, hồ sơ bệnh nhân

Kế toán

Hóa đơn, đơn hàng, phiếu nộp tiền

Quản lý nhân sự

Đơn xin việc, sơ yếu lý lịch

Tài chính và Ngân hàng

Báo cáo tài chính, hồ sơ vay vốn, tờ khai, hợp đồng

Bảo hiểm

Hồ sơ khai báo bảo hiểm, hợp đồng bảo hiểm

Giao vận

Chứng từ, hóa đơn vận chuyển, biên lai

Giáo dục

Bài thi trắc nghiệm, bảng điểm

Chính phủ

Hồ sơ thuế, hợp đồng, phiếu bầu cử

Kiến trúc



Mô hình độc lập và phân tán

Tùy thuộc vào nhu cầu của từng dự án, ABBYY FlexiCapture 10 có thể được triển khai theo hai mô hình: Độc lập và Phân tán.

Mô hình Độc lập (Standalone) được triển khai trên một máy tính, phù hợp với các dự án vừa và nhỏ có số lượng tài liệu cần xử lý không lớn, không thường xuyên và ít người tham gia.



Mô hình Phân tán (Distributed) được triển khai trên nhiều máy chủ chuyên dụng để xử lý dữ liệu (với tính năng cân bằng tải giúp đảm bảo tối đa hiệu suất và hỗ trợ cluster để ngăn ngừa lỗi hệ thống) và các máy trạm cho các khâu trong quá trình xử lý dữ liệu: cấu hình, quản trị, quét ảnh, soát lỗi và giám sát. Phiên bản chạy trên nền web của các trạm quét ảnh và soát lỗi có thể cài đặt một cách dễ dàng.

Mô hình Phân tán phù hợp với các dự án xử lý dữ liệu của doanh nghiệp và chính phủ với quy mô lớn, nhiều người tham gia vận hành.

Kiến trúc máy chủ

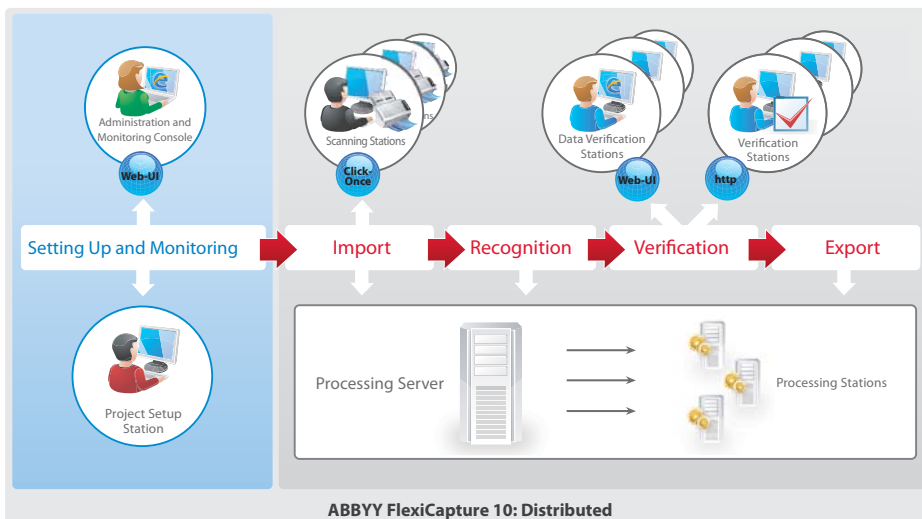
- Máy chủ Ứng dụng (Application Server) xử lý tất cả các tác vụ trong hàng đợi và tự động phân phối các tác vụ này đến các máy trạm theo độ ưu tiên.
- Máy chủ Bản quyền (Licensing Server) đảm bảo các máy trạm và máy chủ hoạt động theo đúng phạm vi của giấy phép sử dụng.
- Máy chủ Dữ liệu (Database Server) lưu trữ thông tin về người sử dụng, lỗi, hoạt động xử lý, các thống kê và dữ liệu được bóc tách.
- Máy chủ Xử lý (Processing Server) đảm nhiệm tất cả các hoạt động cần nhiều tài nguyên như tải ảnh, nhận dạng (phân loại, bóc tách dữ liệu, thực hiện soát lỗi) và kết xuất dữ liệu và ảnh. Máy chủ Xử lý có kiến trúc khả mở để có thể dễ dàng tăng công suất xử lý chỉ bằng cách thêm các Máy trạm Xử lý (Processing Station). Hơn nữa, Máy chủ Xử lý cũng thực hiện cân bằng tải thông qua việc phân bổ hợp lý các tác vụ tới các Máy trạm Xử lý.

Các máy trạm chuyên dụng

- Trạm cấu hình project (Project Setup Station) thực hiện tạo mới, cấu hình, điều chỉnh project (cấu hình tải ảnh, tạo định nghĩa tài liệu, cấu hình luồng tác vụ, loại lỗi...)
- Trạm quét (Scanning Station) thực hiện quét tài liệu, phân tách tài liệu, biên tập sau khi quét. Trạm quét từ xa thực hiện quét và gửi tài liệu theo giao thức HTTP trực tiếp tới FlexiCapture từ bất cứ điểm từ xa nào.
- Công nghệ Microsoft ClickOne cho phép Trạm quét có thể dễ dàng được cài đặt và chạy chỉ bằng việc nhấn vào đường dẫn trên trang web.
- Trạm soát lỗi dữ liệu (Data Verification Station) soát lỗi ở mức thấp - kiểm tra tính chính xác của các ký tự được nhận dạng.
- Trạm soát lỗi dữ liệu trên nền web (Web Data Verification Station) cung cấp giao diện trên nền web để soát lỗi dữ liệu cho phép việc soát lỗi có thể thực hiện từ bất kỳ đâu, bất kỳ lúc nào bằng cách sử dụng một trình duyệt web.
- Trạm soát lỗi (Verification Station) dùng để soát lỗi ở mức cao: kiểm tra các ràng buộc logic, xử lý những trường hợp đặc biệt, xác nhận kết xuất dữ liệu.
- Trạm soát lỗi dữ liệu từ xa (Remote Verification Station) có thể kết nối tới Máy chủ Xử lý theo giao thức HTTP

Công cụ quản trị và cấu hình

- ABBYY FlexiCapture cung cấp công cụ quản trị và giám sát chạy trên nền web với các báo cáo nhằm mục đích giám sát hiệu năng của hệ thống, kiểm soát hoạt động của người dùng.
- Với Crystal Reports Editor, hệ thống cung cấp các công cụ để người quản trị có thể tạo các báo cáo với các tham số và dữ liệu từ cơ sở dữ liệu.





Chức năng chính

Chuẩn bị, Cấu hình và Quản trị

Trong ABBYY FlexiCapture 10, quá trình bóc tách dữ liệu gồm nhiều bước và được chia vào 2 chế độ sử dụng: **chế độ quản trị** (thiết kế mẫu, thiết lập toàn bộ quá trình xử lý) và **chế độ vận hành** (chỉ để thực thi việc xử lý).

ABBYY FlexiCapture cung cấp các công cụ mạnh, dễ sử dụng để thiết kế và tạo các mẫu:

FormDesigner: là công cụ thiết kế các mẫu cố định, biểu mẫu này sau đó sẽ được in và điền bằng tay.

FlexiLayout Studio: là công cụ mạnh và thuận tiện để thiết kế các biểu mẫu cho các tài liệu bán cấu trúc như hóa đơn, hợp đồng.

Document Definition Editor: là công cụ để thiết lập các quy tắc, quy trình xử lý cho mọi loại tài liệu.

Xử lý

Quét và nhập ảnh

Trong trường hợp số lượng tài liệu cần xử lý nhỏ, việc quét ảnh có thể được thực hiện trực tiếp từ FlexiCapture Standalone.

Trong trường hợp số lượng tài liệu lớn, cần kiểm soát chất lượng ảnh quét và thứ tự sắp xếp của tài liệu, nên sử dụng phần mềm chuyên dụng Scanning Station của FlexiCapture Distributed.

Phần mềm có thể được cài đặt hoàn toàn tự động, chỉ bằng thao tác nhấp chuột trên trang web.

Tài liệu cũng có thể gửi đi xử lý thông qua gửi email.

Phân tách và phân loại tài liệu

Để tăng tốc quá trình đối sánh biểu mẫu và bóc tách dữ liệu, ABBYY FlexiCapture hỗ trợ phân tách tài liệu dựa trên các dấu hiệu như trang trắng hay mã vạch, sau đó sắp xếp các tài liệu theo từng loại. Trong ABBYY FlexiCapture 10, người dùng không cần phải tự định nghĩa điều kiện phân loại cho tài liệu, việc định nghĩa này có thể thực hiện tự động dựa trên các ảnh mẫu.

Nhận dạng chữ in, chữ viết tay, đánh dấu và mã vạch

Sở hữu công nghệ nhận dạng vượt trội, ABBYY FlexiCapture nhận dạng với độ chính xác cao chữ in và chữ viết tay của nhiều ngôn ngữ. Công nghệ cũng nhận dạng chính xác mã vạch 1 chiều (barcode), 2 chiều và các đánh dấu (check mark). Chức năng tiến xử lý ảnh mạnh mẽ cũng góp phần nâng cao chất lượng nhận dạng.

Tự động kiểm tra tính hợp lệ của dữ liệu

Các quy tắc kiểm tra tích hợp sẵn trong phần mềm được tự động thực hiện trong quá trình nhận dạng để tăng độ chính xác. Tự động kiểm tra tính hợp lệ của dữ liệu đảm bảo chất lượng nhận dạng cao và giảm bớt công việc cho người soát lỗi.

Soát lỗi dữ liệu

Soát lỗi dữ liệu nằm trong FlexiCapture Standalone và được cài đặt như một trạm riêng biệt trong hệ thống FlexiCapture Distributed. Soát lỗi dữ liệu được sử dụng để soát lỗi ở mức thấp: kiểm tra tính chính xác của dữ liệu nhận dạng so với thông tin trên tài liệu gốc.

ABBYY FlexiCapture cũng cung cấp một ứng dụng chạy trên nền web để soát lỗi dữ liệu. Với ứng dụng này, việc soát lỗi dữ liệu có thể thực hiện ở bất cứ đâu và bất cứ lúc nào bằng cách sử dụng một trình duyệt web. Trạm soát lỗi dữ liệu trên nền web không đòi hỏi phải cài thêm bất cứ chương trình nào. Để bắt đầu xử lý tài liệu, người sử dụng chỉ cần truy cập vào trang web và đăng nhập. Với tính năng mới này, việc phân bổ công việc trở nên dễ dàng hơn, người soát lỗi có thể làm việc ở nhà, trên đường di chuyển hoặc có thể ở các khu vực có chi phí nhân công rẻ. Những trường dữ liệu đặc biệt quan trọng có thể phân bổ cho hai người soát lỗi độc lập để loại bỏ khả năng sai sót. Dữ liệu sau khi soát lỗi sẽ chỉ được chấp nhận khi kết quả soát lỗi của hai người là giống nhau.

Soát lỗi ở mức tài liệu

Soát lỗi tài liệu bao gồm soát lỗi dữ liệu, được thực hiện như một trạm soát lỗi độc lập trong FlexiCapture Distributed và bao gồm tất cả các chức năng soát lỗi. Người soát lỗi cao cấp có thể kiểm tra các lỗi quy tắc và lỗi cấu trúc tài liệu. Họ có thể thay đổi lại cấu trúc tài liệu đối với các tài liệu có lỗi phân loại và cấu trúc tài liệu nhờ thao tác bằng tay trên giao diện "kéo - thả" trực quan. Các tài liệu bị lỗi có thể được gửi lại các Trạm soát lỗi dữ liệu để soát lỗi lại. Cửa sổ tài liệu cho phép xem ảnh ở kích thước gốc của ảnh.

Kết xuất dữ liệu

Dữ liệu có thể được kết xuất tới các ứng dụng hoặc cơ sở dữ liệu hoặc được lưu trong kho lưu trữ của doanh nghiệp. Chương trình có chức năng kết xuất tự động và kết xuất tới nhiều đích khác nhau. ABBYY FlexiCapture hỗ trợ các kiểu kết xuất sau:

- hầu hết các định dạng file phổ biến
- XML cơ bản và nâng cao
- các cơ sở dữ liệu bên ngoài thông qua ODBC
- máy chủ Microsoft Office SharePoint®
- ERP, DMS, CMS và các ứng dụng nghiệp vụ khác thông qua các mô đun tùy biến, các kết nối hoặc script

Biên tập và lưu trữ

Chức năng biên tập cho phép loại bỏ dữ liệu mật hoặc nhạy cảm của một số trường dữ liệu nhất định. Ảnh tài liệu có thể chuyển đổi sang các định dạng tiêu chuẩn trong lưu trữ như PDF hoặc PDF/A.



Đặc tả kỹ thuật

Tính năng đặc biệt

Công nghệ học tự động (Auto-learning technology) làm đơn giản hóa quá trình phân loại và thiết kế biểu mẫu. Với tính năng mới này, người thiết kế chỉ cần thao tác "chỉ và chọn" là có thể mô tả bố cục, cấu trúc của biểu mẫu.

Tính năng nâng cao

Bóc tách dữ liệu từ thư điện tử: Dựa vào khả năng tích hợp với Microsoft Exchange Server và POP3, hệ thống có thể xử lý các ảnh đính kèm trong thư điện tử. Người sử dụng có thể gửi các tài liệu cần xử lý tới địa chỉ thư điện tử quy định trước.

Soát lỗi kép: Những trường dữ liệu đặc biệt quan trọng có thể phân bổ cho hai người soát lỗi độc lập để loại bỏ khả năng sai sót. Dữ liệu sau khi soát lỗi sẽ chỉ được chấp nhận khi kết quả soát lỗi của hai người là giống nhau.

Công cụ tùy biến và tích hợp

Tùy biến các bước và các quy tắc phân luồng tài liệu bằng script. Luồng công việc có thể được mở rộng để phù hợp với yêu cầu như nâng cao chất lượng ảnh thông qua engine của bên thứ ba, kết nối tới các engine nhận dạng (OCR/ICR) khác, thêm các bước soát lỗi mới, ...

Tùy biến giai đoạn quét và soát lỗi bằng script. Script cho phép thêm các chức năng, thay đổi các menu và toolbar trên giao diện quét và soát lỗi.

Processing Script cho phép tùy biến các bước trong quá trình xử lý tài liệu.

Web Service API cho phép truy cập vào Máy chủ xử lý FlexiCapture từ các ứng dụng bên ngoài.

Bằng việc sử dụng API, FlexiCapture có thể tích hợp với nhiều ứng dụng nghiệp vụ hay luồng công việc khác với vai trò là một dịch vụ tự động phân loại tài liệu và bóc tách dữ liệu.

Yêu cầu hệ thống

Bộ xử lý Intel® Core™2/Quad/Pentium®/Celeron®/Xeon™, AMD 6/Turion™/Athlon™/Duron™/Sempron™ hoặc tương thích với 2GHz

Hệ điều hành:

- Máy trạm (độc lập hoặc phân tán): Windows XP SP3, Windows Vista SP2, Windows 7
- Máy chủ (phân tán): Windows Server 2003 SP2, Windows Server 2008 SP2, Windows Server 2008 R2

Đối với các chương trình chạy với giao diện được bản địa hóa, hệ điều hành phải hỗ trợ ngôn ngữ đó.

RAM:

- Máy trạm (độc lập và phân tán): 512 Mb cho mỗi nhân CPU, và tổng cộng không ít hơn 1Gb
- Máy chủ (phân tán): 2GB RAM, 4 GB cho các xử lý cường độ lớn, 2GB cho máy chủ MS SQL

Dung lượng trống của ổ đĩa: Cài đặt độc lập: 1Gb (gồm 700 MB cho cài đặt); Cài đặt phân tán/Máy trạm: 1Gb (gồm 800 Mb cho cài đặt); Cài đặt phân tán/Máy chủ: 100 Mb cho cài đặt và 2Gb cho cơ sở dữ liệu SQL

Máy quét hỗ trợ chuẩn TWAIN, WIA hoặc ISIS

Nhận dạng

Chữ in: 198 ngôn ngữ bao gồm tiếng Việt, tiếng Trung, tiếng Nhật, tiếng Hàn

Chữ viết tay: hơn 100 ngôn ngữ

Đánh dấu: khoanh tròn lựa chọn, đánh dấu vào ô, đánh dấu không có ô, tô kín lựa chọn

Mã vạch: 1D: Code 39, Check Code 39, Interleaved 25, Check, Interleaved 25, EAN 13, EAN 8, Code 128, Codabar, Code 93, IATA 25, UCC-128, UPC-A, UPC-E, Matrix 2/5, Industrial 2/5, PostNet và 2D: PDF 417, Aztec, DataMatrix

Định dạng ảnh đầu vào

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

Định dạng dữ liệu kết xuất

TXT, XLS, DBF, XML, CSV

Định dạng ảnh kết xuất

TIFF, JPEG, JPEG 2000, PCX, BMP, PNG, DCX, PDF, PDF/A (chỉ ảnh, text dưới ảnh)

Dùng thử và Giấy phép sử dụng

ABBYY cung cấp phiên bản dùng thử có đầy đủ chức năng với giới hạn về thời gian và số trang xử lý, cho phép các tổ chức kiểm thử nhiều chức năng của FlexiCapture. Các gói nâng cấp và bảo trì đảm bảo phần mềm cập nhật với công nghệ mới nhất của ABBYY.

Mọi chi tiết xin vui lòng liên hệ Đối tác chiến lược và Nhà phân phối độc quyền của ABBYY tại Việt Nam



Công ty Đầu tư và Phát triển Đông Kinh

Tầng 6, tòa nhà Intracom, đường Dịch Vọng Hậu, Cầu Giấy, Hà Nội
(+ 844) 2219 6396 info@dkc.com.vn www.dkc.com.vn

www.SOHOA.com.vn

ABBYY 3A

Asia, Africa, Baltic States, Middle East, South America

P.O. Box #20, Moscow, Russia, 127273

Tel: +7 495 783 3700,

Fax: +7 495 783 2663,

sales_3A@abbyy.com

www.ABBYY.com

ABBYY®