

ABBYY® Recognition Server 3.0



Giải pháp số hoá tài liệu thông minh

ABBYY Recognition Server 3.0 là giải pháp số hoá tài liệu mạnh mẽ và mang lại hiệu quả kinh tế cao. Với khả năng số hoá, chuyển đổi tài liệu giấy về tài liệu điện tử - là các định dạng tài liệu có thể biên tập và tìm kiếm được, lập chỉ mục và đưa các tài liệu này vào các hệ thống lưu trữ điện tử, hệ thống quản lý nội dung hoặc các hệ thống tìm kiếm, ABBYY Recognition Server 3.0 góp phần cải tiến, chuẩn hóa quy trình nghiệp vụ cho mọi doanh nghiệp, tổ chức.

Công nghệ tiên tiến

Sở hữu công nghệ OCR tiên tiến có khả năng nhận dạng gần 200 ngôn ngữ, ABBYY Recognition Server cho kết quả nhận dạng với độ chính xác rất cao. ABBYY Recognition Server cho phép chuyển đổi ảnh quét tài liệu sang nhiều định dạng tài liệu điện tử, đáp ứng được mọi nhu cầu lưu trữ, chia sẻ thông tin và biên tập tài liệu, ví dụ như các định dạng của Microsoft® Office 2007 hay các định dạng PDF, PDF/A, XML, RTF.

Được thiết kế để xử lý số lượng tài liệu lớn

Với ABBYY Recognition Server, việc mở rộng quy mô xử lý là hết sức dễ dàng. Công suất xử lý của hệ thống có thể được nâng cao nhanh chóng bằng cách đơn giản là thêm CPU hay máy chủ. Hệ thống có thể hoạt động theo lịch trình đặt trước hoặc liên tục 24/7.

Triển khai nhanh chóng, sử dụng dễ dàng

ABBYY Recognition Server có giao diện người dùng thân thiện và trực quan, cài đặt đơn giản và sử dụng dễ dàng nên không đòi hỏi người dùng phải có đào tạo trước đó. Để vận hành hệ thống, người dùng cũng chỉ cần những thao tác tối thiểu bởi các khâu như nhận dạng hay chuyển đổi tài liệu được thực hiện hoàn toàn tự động. Các thao tác cần làm bằng tay như quét tài liệu, lập chỉ mục và soát lỗi cũng được ABBYY Recognition Server hỗ trợ bằng nhiều công cụ thuận tiện.

Tương thích với các loại máy quét và hệ thống quản lý nội dung

ABBYY Recognition Server được thiết kế để sẵn sàng là phần hệ nhận dạng bổ sung cho các thiết bị quét hoặc các phần mềm khác. ABBYY Recognition Server có thể tích hợp với máy quét hay ứng dụng khác bằng nhiều cách, từ phương thức trao chuyển dữ liệu qua e-mail cho đến tích hợp chặt chẽ qua gọi hàm API.

Đấu nối với các máy tìm kiếm của Microsoft và Google

ABBYY Recognition Server không những có thể được dùng như là một giải pháp số hoá tài liệu độc lập mà nó còn có thể hoạt động như là một nền máy chủ nhận dạng cho các máy tìm kiếm như Google Search Appliance™, Microsoft Office SharePoint® Server hay Windows® Desktop Search bằng cách cho phép các hệ thống kể trên thực hiện việc lập chỉ mục và tìm kiếm trên nội dung của tài liệu dạng ảnh.

Ngôn ngữ mới

Một số ngôn ngữ mới được công nghệ nhận dạng của ABBYY Recognition Server 3.0 hỗ trợ: **tiếng Việt**, tiếng Trung (giản thể và phồn thể), tiếng Nhật, tiếng Hàn và tiếng Yiddish.

Lợi ích của sản phẩm

Đảm nhiệm toàn bộ quá trình số hoá tài liệu

ABBYY Recognition Server 3.0

thực hiện tất cả các khâu của quá trình số hoá tài liệu:

- Quét tài liệu
- Lập chỉ mục
- Kết xuất tài liệu ra các hệ thống lưu trữ và quản lý nội dung

Quy tắc phân tách tài liệu

ABBYY Recognition Server cho phép người quản trị sử dụng các mã script để thiết lập các quy tắc phân tách tài liệu linh hoạt.

Nhanh chóng thu hồi vốn đầu tư

ABBYY Recognition Server là phần mềm đem lại hiệu quả kinh tế cao, dễ cấu hình và tích hợp vào luồng công việc sẵn có. Sản phẩm có thể được cài đặt chỉ trong vài giờ đồng hồ và ngay lập tức tham gia vào quá trình số hoá tài liệu trong tổ chức.

Chức năng



Thu nhận ảnh đầu vào

Máy chủ quản lý (Server Manager) thu nhận ảnh từ các nguồn đầu vào và đưa chúng vào hàng đợi chờ xử lý.

Các nguồn đầu vào có thể là:

- thư mục chia sẻ trên mạng nội bộ;
- thư mục FTP;
- thư mục trong hộp thư;
- trạm quét.

Nhận dạng

Các ảnh trong hàng đợi lần lượt được chuyển tới các Trạm xử lý (Processing Stations) để nhận dạng. Nếu có nhiều Trạm xử lý trong hệ thống thì Máy chủ quản lý sẽ tự động cân bằng tải xử lý của các trạm này.

Soát lỗi (tùy chọn)

Nếu có yêu cầu soát lỗi thì các trang tài liệu cần soát lỗi sẽ được xếp vào hàng đợi. Máy chủ quản lý sẽ tự động chuyển các trang này tới các Trạm soát lỗi.

Phân tách tài liệu

Các trang tài liệu trong lô được gom thành các tài liệu dựa trên các dấu hiệu phân tách tài liệu như: trang trắng, mã vạch được in (dán) trên trang đầu tiên của mỗi tài liệu. Việc phân tách tài liệu còn có thể được thực hiện bằng cách đặt ra các quy tắc tự động phân tách dựa trên dữ liệu nhận dạng được, các quy tắc này được định nghĩa bằng các mã lệnh script.

Lập chỉ mục (tùy chọn)

Nếu có yêu cầu đánh chỉ mục, người vận hành chỉ cần khoanh vùng các phần cần thiết trong trang, giá trị của các trường thuộc tính sẽ được điền tự động. Thêm vào đó, có thể viết mã lệnh script để tự động đoán nhận kiểu tài liệu và đánh chỉ mục tài liệu.

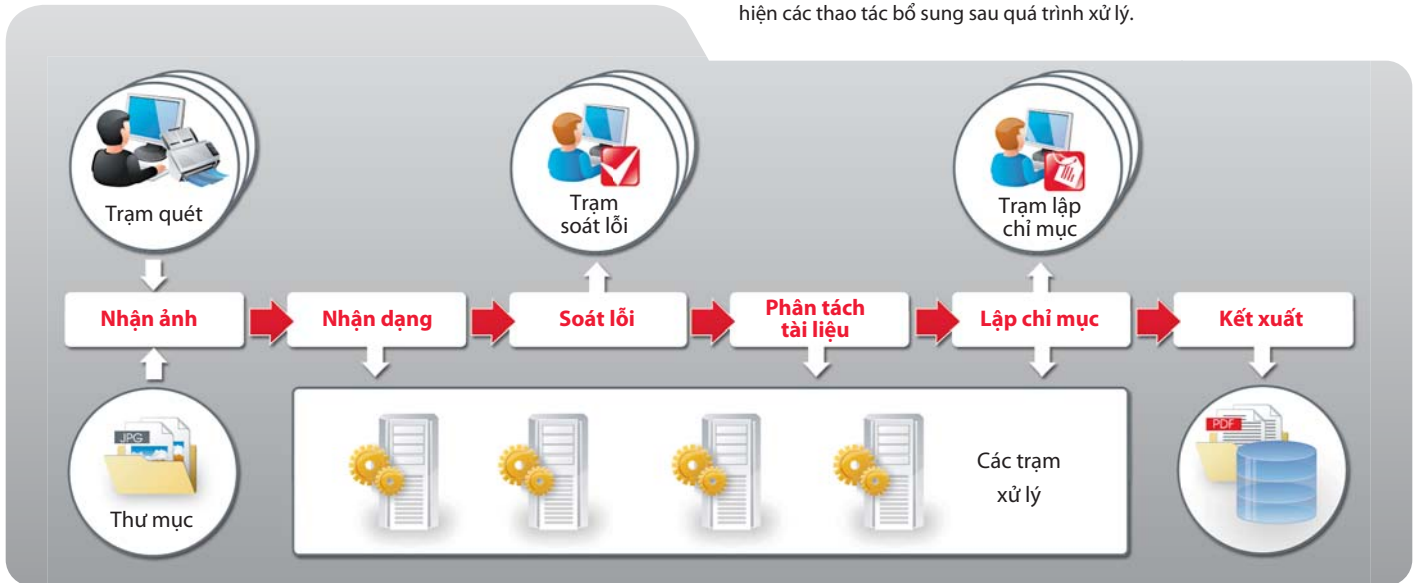
Kết xuất dữ liệu

Sau khi tất cả các trang tài liệu được nhận dạng, Máy chủ quản lý sẽ chuyển tài liệu đầu ra đến nơi cần kết xuất.

Nơi kết xuất dữ liệu có thể là:

- thư mục trên chia sẻ trên mạng nội bộ;
- thư viện SharePoint;
- địa chỉ e-mail.

Hơn nữa, có thể viết mã lệnh script để xử lý kết quả kết xuất cũng như thực hiện các thao tác bổ sung sau quá trình xử lý.



Khả năng chịu lỗi

ABBYY Recognition Server 3.0 được thiết kế để có thể hoạt động mà không cần có người giám sát, phần mềm có những tính năng chịu lỗi đặc biệt để đảm bảo tính bền bỉ của hệ thống:

- Các thành phần chính như Máy chủ quản lý và các Trạm xử lý là các Windows Services, chúng tự động hoạt động khi Windows khởi động.
- Máy chủ quản lý theo định kỳ kiểm tra các kết nối với các trạm làm việc. Nếu phát hiện có kết nối bị mất, Máy chủ quản lý sẽ tự động thiết lập lại kết nối đó.
- Quy trình xử lý tài liệu đảm bảo không dữ liệu nào bị mất ngay cả khi xảy ra lỗi hệ thống. Nếu một trạm đang xử lý tài liệu gặp lỗi thì tài liệu đang được xử lý đó sẽ được tự động chuyển sang một trạm khác đang rảnh hoặc nếu tất cả các trạm đều bận, tài liệu sẽ được đưa vào hàng đợi chờ cho đến khi có trạm rảnh.

Xử lý lỗi

- Quản lý chất lượng - Người quản trị có thể đặt ngưỡng cho độ tin cậy về chất lượng nhận dạng của các tài liệu. Khi đó, các tài liệu có chất lượng nhận dạng kém sẽ không được kết xuất ra nơi lưu trữ đã chỉ định, thay vào đó chúng sẽ được lưu trữ tại một thư mục riêng để có những xử lý cần thiết sau đó.
- Huỷ tiến trình xử lý - Người quản trị có thể đặt ngưỡng cho thời gian tự động huỷ tiến trình xử lý. Khi đó, hệ thống sẽ tự động huỷ các tiến trình có thời gian hoạt động vượt ngưỡng.
- Nhật ký - Hệ thống ghi nhật ký mọi sự kiện xảy ra. Nhật ký hệ thống này giúp người quản trị trong quá trình dò lỗi.

Kiến trúc

ABBYY Recognition Server gồm những thành phần chính sau đây:

Máy chủ quản lý (Server Manager) - là thành phần chính của Recognition Server, hoạt động dưới dạng Windows service, điều phối hoạt động của các thành phần khác trong hệ thống, quản lý bản quyền và tắt cả các lựa chọn khi xử lý tài liệu, phân phối các lệnh nhận dạng và chuyển định dạng file giữa các trạm xử lý, soát lỗi và lập chỉ mục.

Trạm xử lý (Processing Station) - hoạt động dưới dạng Windows service, thực hiện việc nhận dạng chữ (OCR) và chuyển đổi định dạng tài liệu. Quy mô xử lý của hệ thống có thể được mở rộng bằng cách thêm các Trạm xử lý.

Trạm quét (Scanning Station) - là phần mềm quét tài liệu có giao diện sử dụng thân thiện và trực quan.

Các tính năng chính của ABBYY Scanning Station:

- Khởi động và dừng quá trình quét, thiết lập thông số quét;
- Xem và quản lý chất lượng ảnh đã quét;
- Phân tách lô các ảnh quét thành các tài liệu riêng rẽ dựa trên các dấu hiệu phân tách.

ABBYY Recognition Server 3.0 Scanning Station hỗ trợ các trình điều khiển máy quét TWAIN, WIA và ISIS.

Trạm soát lỗi (Verification Station) - Trạm xử lý cung cấp công cụ soát lỗi kết quả nhận dạng.

Trạm lập chỉ mục (Indexing Station) - cung cấp chức năng thiết lập kiểu tài liệu và các công cụ để nhập giá trị cho các thuộc tính của tài liệu. Với mỗi tài liệu, người vận hành chọn một kiểu tài liệu thích hợp rồi điền giá trị cho các trường thuộc tính tương ứng với kiểu tài liệu đã chọn. Tài liệu ở trạm lập chỉ mục là tài liệu đã nhận dạng.

Console quản trị từ xa - là một Microsoft Management Console snap-in, cung cấp giao diện để cấu hình và giám sát hoạt động của Recognition Server.

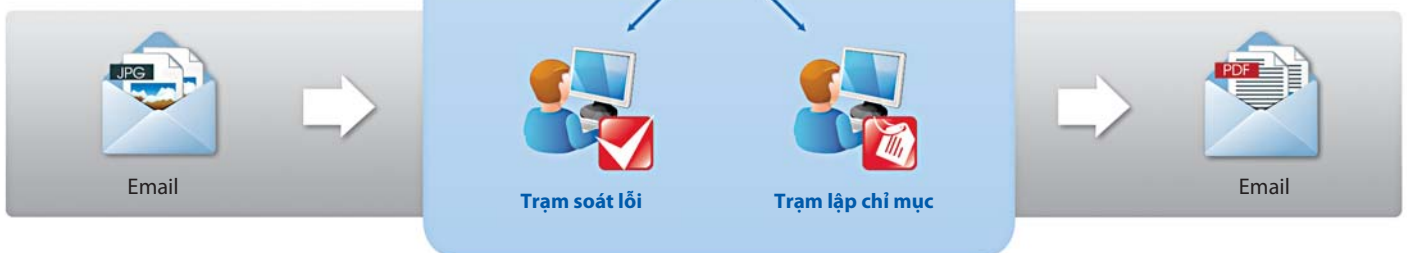
Luồng 1



Luồng 2



Luồng 3



Tích hợp

ABBYY Recognition Server 3.0 cung cấp nhiều cách tích hợp với các hệ thống khác, cũng như với các máy quét và máy đa chức năng (Multifunction Printer):

- COM-based API - là một giao diện lập trình theo chuẩn COM, cho phép Recognition Server trao chuyển các file ảnh và kết quả nhận dạng với các ứng dụng khác.

- Web Service API - là giao diện tích hợp qua Web Service. Giao diện này cho phép tích hợp Recognition Server với các hệ thống khác qua giao thức HTTP cũng như sử dụng Recognition Server trong một môi trường hướng dịch vụ (SOA).
- XML ticket - các tham số nhận dạng có thể được thiết lập thông qua file XML (XML ticket), các file XML này có thể được sinh ra tự động từ chương trình.

ABBYY Recognition Server 3.0



Các thông số kỹ thuật

Yêu cầu hệ thống

- PC với bộ xử lý 500 MHz hoặc cao hơn
- Hệ điều hành: Microsoft Windows Server 2008, Windows Vista®, Windows Server 2003, Windows XP, Windows 2000
- RAM 128 MB và 100 MB thêm cho mỗi tiến trình nhận dạng
- Ổ đĩa cứng 700 MB
- Phiên bản Microsoft.NET Framework 1.1 hoặc mới hơn
- Microsoft Outlook 2000 hoặc mới hơn (để xử lý và gửi e-mail)

Ngôn ngữ nhận dạng

- 37 ngôn ngữ chính với ký tự La-tinh, Kirin, Ác-mê-ni-a cùng từ điển hỗ trợ
- 133 ngôn ngữ khác với ký tự La-tinh, Kirin hoặc Hy Lạp
- Tiếng Việt, tiếng Hê-brơ cùng từ điển hỗ trợ
- Tiếng Trung, tiếng Nhật, tiếng Hàn và tiếng Thái
- 5 ngôn ngữ châu Âu cổ
- 4 ngôn ngữ nhân tạo (Esperanto, Interlingua, Ido, Occidental)
- 6 ngôn ngữ lập trình
- Các công thức hoá học đơn giản
- Ký tự số

Kiểu in

- Máy in thông dụng, máy đánh chữ, máy in kim, in Gô-tích, OCR-A, OCR-B, MICR (E13B), máy Fax
- Kiểu mã vạch một chiều: Check Code 39, Check Interleaved 25, Code 128, Code 39, EAN 13, EAN 8, Interleaved 25, CODABAR, UCC Code 128, Code 2 of 5 (Industrial, IATA, Matrix), Code 93, UPC-A, UPC-E và Postnet
- Mã vạch hai chiều: PDF 417, Aztec, DataMatrix, QRcode

Các định dạng ảnh đầu vào

- TIFF: đen trắng, xám, màu, nhiều trang
- Phương thức nén: Unpacked, CCITT Group 3, CCITT Group 3 FAX(2D), CCITT Group4, PackBits, JPEG, ZIP, LZW
- JPEG, JPEG 2000 part1: xám, màu
- PDF
- DjVu
- BMP: đen trắng, xám, màu
- PNG: đen trắng, xám, màu
- PCX, DCX: đen trắng, xám, màu

Các định dạng đầu ra

- PDF, PDF/A-1a, PDF/A-1b
- RTF
- DOC, DOCX, Word XML
- XLS, XLSX
- TXT, CSV
- HTML
- TIFF
- JPEG, JPEG 2000
- XML
- Định dạng riêng của FineReader (tương thích với FineReader Engine)

Các ứng dụng chính

Xây dựng các hệ thống lưu trữ, bộ sưu tập và các thư viện số

ABBYY Recognition Server đảm nhiệm toàn bộ quá trình số hoá tài liệu, cung cấp các công cụ tiện ích cho việc quét tài liệu, nhận dạng, soát lỗi, thiết lập thuộc tính cho tài liệu, đánh chỉ mục toàn văn và chuyển đổi tài liệu. Phần mềm chuyển các tài liệu quét sang định dạng PDF và PDF/A - là những định dạng thích hợp cho công tác lưu trữ, bảo quản tài liệu dài hạn. ABBYY Recognition Server được thiết kế để vận hành 24/7 và có thể giải quyết thành công các dự án lớn nhất, nơi mà hàng triệu trang tài liệu được số hóa trong thời gian rất ngắn.

Dây chuyền số hóa tài liệu cho các hệ thống ECM

Với các trạm xử lý mới như Scanning Station và Indexing Station, ABBYY Recognition Server giờ đây đã cung cấp một dây chuyền số hóa hoàn chỉnh, đáp ứng hoàn toàn các yêu cầu số hoá tài liệu của các hệ thống CMS/DMS/ECM. Tài liệu được phân loại tự động hoặc bởi người vận hành, sau đó dựa trên loại và các đặc tính của tài liệu mà nó được đưa đến một nơi thích hợp. ABBYY Recognition Server cũng có thể hoạt động như một dịch vụ số hoá dạng web service, nó cho phép tải tài liệu lên từ xa, do đó việc khai thác sản phẩm này không bị giới hạn lại trong một văn phòng mà còn mở rộng được ra cả ở bên ngoài.

Phân hệ nhận dạng cho các máy tìm kiếm

ABBYY Recognition Server có thể được sử dụng như một phân hệ nhận dạng cho các máy tìm kiếm của Google và Microsoft. Thông thường, các máy tìm kiếm chỉ có thể lập chỉ mục và tìm các tài liệu chứa text như: HTML, DOC, PDF, v.v... Nhưng rõ ràng là không phải mọi thông tin trong doanh nghiệp đều được lưu dưới các định dạng thích hợp cho việc tìm kiếm toàn văn. Một số thông tin có thể được lưu dưới dạng tài liệu quét, bản fax và những thông tin này là không thể tìm kiếm được. ABBYY Recognition Server giúp tự động chuyển đổi các tài liệu này sang các định dạng cho phép tìm kiếm được.

Dịch vụ nhận dạng

Với những tổ chức có hàng trăm, thậm chí hàng ngàn nhân viên, ABBYY Recognition Server là một hệ thống nhận dạng đáp ứng nhiều lệnh nhận dạng và chuyển đổi file, có thể truy cập thông qua e-mail và/hoặc cấu hình trong các máy in đa chức năng của công ty. Nhân viên có thể sử dụng dịch vụ nhận dạng tại bất kỳ thời điểm nào, kể cả ngoài giờ làm việc, từ mọi địa điểm. Hơn nữa, người quản trị không cần cài đặt chương trình vào mỗi trạm làm việc trong mạng nội bộ mà chỉ cần cài đặt, cấu hình hệ thống trung tâm một lần.

Dùng thử và Bản quyền sử dụng

ABBYY có cơ chế cấp bản quyền sử dụng linh hoạt, nhờ đó khách hàng có thể lựa chọn bản quyền phù hợp nhất với nhu cầu xử lý và khả năng đầu tư của mình. Tùy theo khối lượng tài liệu cần xử lý và các chức năng cần có, khách hàng có thể tùy chọn bản quyền sử dụng theo các thông số sau:

- Số trang tài liệu cần xử lý trong 1 năm
- Số nhân CPU và số máy trạm sẽ sử dụng
- Các chức năng cao cấp cần bổ sung

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®