

## ABBYY® Recognition Server 3.5

Professional Edition & Extended Edition



### Eine zuverlässige Lösung für OCR und Dokumentenerfassung

Die serverbasierte Lösung ABBYY Recognition Server ermöglicht Unternehmen eine effiziente Umwandlung von Papierdokumenten oder Dokumentenbildern in durchsuchbare und editierbare elektronische Dateien. Die leistungsstarke Technologie unterstützt das Scannen von Dokumenten, eine präzise Volltext-OCR sowie das Auslesen von Metadaten, um Inhalte für digitale Archive, Enterprise-Content-Management (ECM), Enterprise Search und andere dokumentenbasierte Geschäftsprozesse bereit zu stellen. Eine skalierbare Architektur ermöglicht eine effiziente, automatisierte und zentral verwaltete Erfassung großer Dokumentenmengen, die auch in andere Anwendungen und Backend-Systeme integriert werden kann.

### Wo wird ABBYY Recognition Server eingesetzt?

Der Erfolg eines Unternehmens ist heutzutage stark von einer stetigen Erhöhung der Produktivität und der optimalen Verwertung der Informationen abhängig, mit denen die Organisation täglich aufs Neue überflutet wird. Unter Nutzung der neuesten Technologien automatisiert Recognition Server die Dokumentenverarbeitung in Unternehmen und konvertiert papierbasierte und digitale Informationen in wiederverwendbare Formate wie beispielsweise indexierte PDFs. Die so erstellten Daten sind durchsuchbar und können in weiteren Unternehmensprozessen genutzt werden.

#### • Dokumentenarchivierung und Konvertierung bereits archivierter Dokumente und Daten

Recognition Server ermöglicht die Umwandlung großer Dokumentenmengen in durchsuchbare Volltextformate für Dokumentenmanagement und Archivierung. Recognition Server bietet sowohl eine Volltextumwandlung als auch eine einfache Indexierung bildbasierter Dokumente. Nach der Umwandlung der Dokumente in PDF, XML\* und andere Formate haben die Anwender die Möglichkeit, die Erkennungsqualität zu überprüfen und die Indexangaben manuell hinzuzufügen. Mit den einfach zu nutzenden Integrationstools lässt sich die Dokumentenablage direkt in Archiven, SharePoint Servern oder anderen Backend-Systemen automatisieren.

#### • Dokumentenkonvertierung rund um die Uhr

Als ein serverbasiertes System bietet Recognition Server zeitlich unbegrenzten Zugriff auf OCR-Umwandlung, und dies ohne die Installation von Software auf die einzelnen Workstations. Die Benutzer können einfach eine E-Mail senden oder Dokumente in einem vorgegebenen Ordner ablegen und erhalten so in kürzester Zeit exakt umwandelte und durchsuchbare Dokumente. Recognition Server kann außerdem mit MFPs und Netzwerkscannern verknüpft werden. Auf diese Weise können gescannte oder kopierte Dokumente jeden Tag automatisch verarbeitet werden.

#### • Erhöhte Produktivität für den juristischen Sektor

Recognition Server bietet eine flexible und exakte Technologie, um die wachsende Menge an Informationen zu verwalten, die täglich Anwaltskanzleien erreichen. Auch wenn viele der elektronischen Dokumente in einem durchsuchbaren elektronischen Format zur Verfügung stehen, sind E-Mail-Anhänge wie gescanntes PDF, JPEG oder TIFF weiterhin nicht durchsuchbar, was zu einer Verzögerung bei der Erfassung wichtiger Daten führen kann. Recognition Server wandelt diese „nicht zugänglichen“ Informationen schnell und einfach in durchsuchbare PDFs um - vervollständigt mit Bates Stempeln und in der ursprünglichen Formatierung.

#### • Enterprise Search Systeme

Recognition Server bietet Module für Enterprise Search Systeme. Die Module für Google Search Appliance™ (GSA) und Microsoft® IFilter\*\* unterstützen Unternehmen beim Auslesen von Informationen aus Dokumenten in Bildformaten. Damit können Unternehmen auf Informationen zugreifen, die bereits im Unternehmensnetzwerk vorhanden, jedoch bisher nicht per Suchfunktion auffindbar waren.

### OCR und PDF-Umwandlung über eine serverbasierte Architektur

- **Automatisiert und unabhängig.** Zeitgesteuerte OCR startet sofort nach Eingang der Dokumente in überwachten Ordnern oder zu festgelegten Zeiten. Arbeitet als Hintergrundservice und unabhängig von den Client-Workstations.
- **Mehrere Zugriffspunkte.** Zugriff auf Dokumente über MFPs, Scanner, das Firmennetzwerk, FTP-Ordner und E-Mail-Eingang.
- **Verarbeitung großer Dokumentenmengen.** Maximierung der Kapazität durch parallele Verarbeitung und automatische Verteilung der Arbeitslast.
- **Skalierbarkeit.** Erhöhte Leistung durch das Hinzufügen von Workstations und CPU-Kernen.
- **Zentralisierte Verwaltung.** Zentrale Verwaltung und Administration.
- **Genauigkeit.** Über 20 Jahre an Erfahrung mit Erkennungstechnologien sorgen für eine präzise und genaue Textumwandlung.
- **International.** Erkennung in über 190 Sprachen.
- **Fehlertoleranz.** Ereignisprotokollierung, automatischer Start oder Neustart der OCR-Umwandlung und Umleiten der Dokumente bei Bedarf.
- **Flexible Integration.** XML-Ticket Support, COM-basierte API- und Web Service API ermöglichen die Integration in andere Anwendungen.

\* Export zu XML steht nur in Recognition Server Extended Edition zur Verfügung.

\*\* Anbindung an Google Search Appliance (GSA) und iFilter für Microsoft Windows® Search (Desktop und Server) stehen in Recognition Server Extended Edition zur Verfügung. Sie können als Zusatzmodule für den Recognition Server Professional Edition erworben werden.

# Funktionsüberblick

## Überblick: Schritte der Dokumentenverarbeitung

### 1. Scannen und Importieren von Dokumenten:

- **Scannen:** Bilder werden von einem Operator auf der Scanstation gescannt und dann an Recognition Server gesendet.
- **Importieren von Dokumenten:** Der Server Manager importiert Dateien aus bestimmten Eingabeordnern (gemeinsam genutzte Ordner, FTP-Ordner oder E-Mail Postfächer) und importiert diese in die Verarbeitungswarteschlange.

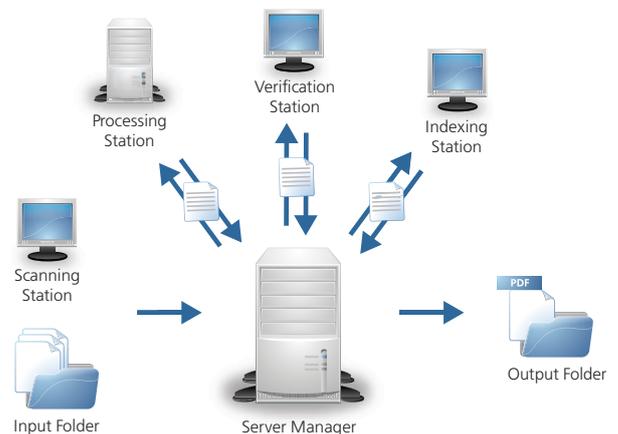
**2. Erkennung:** Recognition Server verteilt die Dateien für die Erkennung auf die verfügbaren Verarbeitungsstationen. Die einzelnen Workflows für die Verarbeitung der Dokumente auf den verfügbaren Verarbeitungsstationen werden vom Server Manager verwaltet und überwacht. Beide Komponenten arbeiten als Dienste unter Windows® und können einzeln im Netzwerk oder zusammen auf einer Workstation installiert werden.

**3. Verifizierung (optional):** Bei Dokumenten mit niedriger Bildqualität oder wenn eine besonders hohe Erkennungsgenauigkeit erforderlich ist, können die jeweiligen Dokumente optional an eine Verifizierungsstation zur manuellen Korrektur durch einen Mitarbeiter gesendet werden. Der Server Manager übergibt die Dokumente entsprechend den Workfloweinstellungen an die Verifizierungsstationen.

**4. Dokumententrennung:** Nach Erkennung und Überprüfung können Dokumente anhand von Leerseiten, Strichcodes oder nach festgelegter Seitenzahl automatisch getrennt werden. Zur Erstellung eigener Trennungsregeln können auch Skripte eingesetzt werden.

**5. Klassifizierung und Indexierung (optional):** Unternehmen können wählen, ob die Daten manuell oder anhand individuell angepasster Skripte indexiert werden sollen. Manuelle Indexierung kann auf der Indexierungsstation durchgeführt werden. Skripte erlauben automatische Indexierung aufgrund von Dokumenteigenschaften und Metadaten.

**6. Exportieren:** Zum Schluss übergibt der Server Manager das Dokument an das vordefinierte Ziel, beispielsweise an einen Netzwerkordner, eine SharePoint®-Bibliothek\*, eine E-Mail-Adresse oder an weitere Anwendungen, die die Recognition Server API nutzen.



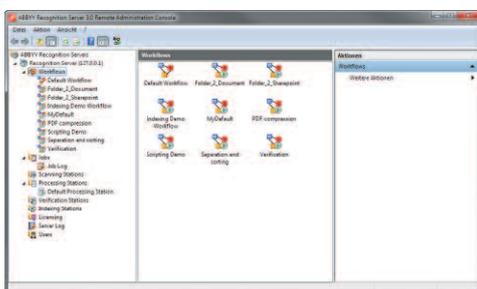
## Konfiguration und Management

### Management und Administration

ABBYY Recognition Server wird per Fernzugriff über die Microsoft Management Console (MMC) verwaltet. Alle Systemeinstellungen einschließlich Workflows, Joblisten, Eigenschaften der Verarbeitungsstationen, Lizenzen und Serverprotokolldateien können von einem zentralen Standort aus erreicht werden.

### Workflows zur Konvertierung und Jobverteilung

Ein Workflow in Recognition Server besteht aus einer Reihe von Verarbeitungsparametern, die Dokumenteneingabequellen, Verarbeitungsstufen und Bereitstellungsmethoden beinhalten. Ein Job ist die kleinste Verarbeitungseinheit und kann ein oder mehrere Bilder oder PDF-Dateien enthalten. Der Administrator legt im Voraus die Prioritäten für jeden Workflow fest und stellt so sicher, dass alle Jobs in der Reihenfolge ihrer Priorität verarbeitet werden.

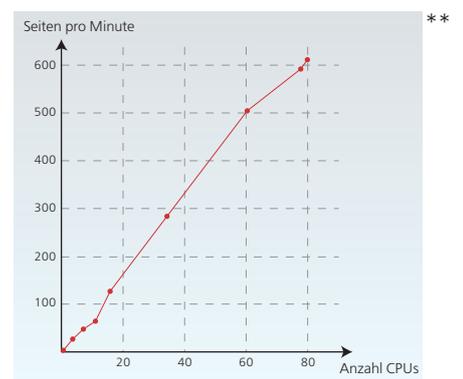


### Zeitgesteuerte und effiziente Verarbeitung

Um die Effizienz zu erhöhen und weniger arbeitsintensive Phasen (z.B. in der Nacht) auszunutzen, können Administratoren einzelne Workflows zu einem bestimmten Zeitpunkt oder in einem Zeitintervall (z.B. täglich, wöchentlich, monatlich) starten lassen. Zudem können sie die verschiedenen Verarbeitungsstationen zu unterschiedlichen Zeiten einsetzen.

### Skalierbarkeit

ABBYY Recognition Server erfüllt verschiedene Anforderungen an die zu verarbeitende Datenmenge. Server Manager und Verarbeitungsstationen können auf einem System mit mehreren CPU-Kernen oder dezentralisiert im Netzwerk installiert werden. Die flexible Architektur lässt sich effizient skalieren und ermöglicht Einrichtung von Systemen, die Hunderte von Seiten pro Minute verarbeiten können.



\* Export zu Microsoft SharePoint ist nur in Recognition Server Extended Edition verfügbar.

\*\* Die Informationen bezüglich Diagramm und Verarbeitungskapazität basieren auf internen Tests. Die Systemleistung ist stark abhängig von der Bildqualität, der Hardwareleistung, der Netzwerkkonfiguration und anderen Faktoren.



# Funktionsüberblick

## Funktionen

### Import und Scannen

Mit ABBYY Recognition Server wird eine Scan-Station zum direkten Scannen mitgeliefert. Sie unterstützt die Schnittstellen TWAIN, WIA und ISIS und beinhaltet Tools für die Vorverarbeitung der Dokumentenbilder vor der eigentlichen Erkennung, wie z.B. Dokumentenvorschau, verschiedene Bildverbesserungsoptionen, Stapeltypdefinitionen, manuelles Ausschwärzen und Seitenreihenfolge. Durch Skripting-Befehle lässt sich der Input-Prozess anpassen, z.B. große Seiten werden automatisch geteilt oder einzelne Seite nach dem Duplex-Scannen neu geordnet.

### Erkennung und PDF-Umwandlung

Recognition Server basiert auf ABBYYs preisgekrönter, für eine hohe Erkennungsgenauigkeit und Stabilität bekannte Erkennungstechnologie, und bietet eine breite Auswahl an Funktionalitäten und Dateiformaten:

- **Bildvorverarbeitung:** Aufteilung von Doppelseiten (z.B. bei Büchern), Konvertierung farbiger und grauer Bilder in Schwarz-Weiß, Entfernung von Bildrauschen und mehr.
- **Drucktyp:** Normaler Text, Schreibmaschine, Punktmatrix, OCR-A, OCR-B und MICR (E13b)
- **Sprachen:** Über 190 OCR-Sprachen, auch mehrsprachige Dokumente werden verarbeitet\*
- **Spezielle Schriftarten:** Unterstützung für Fraktur, Schwabacher und die meisten anderen gotischen Schriftarten welche zwischen 1700 und 1937 in der englischen, deutschen, französischen, italienischen und spanischen Schriftsprache genutzt wurden\*\*
- **Barcodes:** Erkennt viele 1D- und 2D-Barcodes einschließlich der Barcodes 2D Aztec, Data Matrix und QR Code-Barcodes, wobei Anordnung sowie Positionierung auf einem Dokument keine Rolle spielen
- **Verarbeitungsgeschwindigkeit:** Drei verschiedene Modi – Genauigkeit, Geschwindigkeit und Balanced Mode

### Anpassung über Skripte

Die Nutzung der Skripte erlaubt Administratoren, die interne Verarbeitung, Ausnahmeregelungen und den Export der Dokumente anzupassen. So können flexiblere Regeln für den Import oder die Trennung der Dokumente implementiert werden sowie eine automatische Indexierung oder individuelle Exportfunktionen eingeführt werden, die eine Veröffentlichung der Ausgabedokumente zu externen Systemen vereinfachen. Skripte können zusätzlich zur Definition von Ausnahmen verwendet werden, um ausgewählte Dokumente separat zu behandeln.

### Qualitätskontrolle und Verifizierung

Verifizierungsstationen, die auf mehreren Workstations installiert und dank der Concurrent-Lizenz von mehreren Anwendern genutzt werden können, ermöglichen eine manuelle Überprüfung der Ergebnisse. Die Mitarbeiter überprüfen das Layout (Bilder, Texte, Tabellen), bearbeiten den Text und überprüfen die Rechtschreibung.

### Trennung und Benennung von Dokumenten

Recognition Server kann anhand von Leerseiten oder Barcodes die einzelnen Dokumente in einem Stapel voneinander trennen. Außerdem können Dokumente aufgrund einer festgelegten Seitenzahl erstellt werden. Dokumente, die sich in verschiedenen Unterordnern befinden,

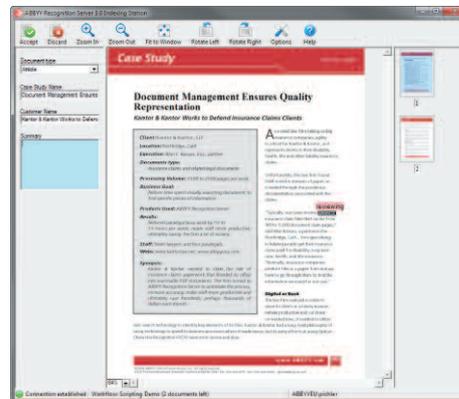
können zu einem Dokument zusammengefasst werden. Anhand von Merkmalen wie z.B. Datum, Zeit, vorgegebenem Text und Barcodewerten können einzelnen Dokumenten automatisch Namen zugefügt werden. Zusätzlich können anhand von Skripts eigene Regeln zur Dokumententrennung erstellt werden.

### Klassifizierung und „Point-and-Click“ Indexierung

Die Indexierungsstation bietet eine Schnittstelle, über die Anwender Dokumententypen und deren Attribute definieren können. Für die Definition der Indizes stehen den Anwendern drei verschiedene Möglichkeiten zur Verfügung.

- **Indexierung per Eingabe:** Mitarbeiter geben die Indizes über die Tastatur manuell ein
- **„Point-and-Click“-Indexierung:** Mitarbeiter extrahieren Schlüsselwörter für Indizes, indem sie den relevanten Text per Mausklick automatisch aus dem Dokumentenbild extrahieren
- **Indexierung von Schlüsselsätzen:** Mitarbeiter nutzen eine sogenannte "Rubberband"-Funktion für das Markieren von Sätzen, mehreren Zeilen oder ganzen Absätzen.

Darüber hinaus können Administratoren mithilfe von selbst entwickelten Skripten automatische Indexierungsfunktionen auf Basis der Dokumentenstruktur und des Inhalts in ihre Workflows einbinden.



### Ausgabeformate

Recognition Server unterstützt eine Vielzahl von Ausgabeformaten einschließlich DOC, DOCX, RTF, XML\*\*\*, XLS, XLSX, HTML, TIFF, JPEG, JPEG 2000 und vielen mehr. Es können mehrere Ausgabeformate erzeugt werden. Recognition Server unterstützt das Exportieren in „einfache“ durchsuchbare PDFs, lineare und Tagged PDFs, PDF-Dateien mit Sicherheitsoptionen und Verschlüsselung sowie PDF/A für die Langzeitarchivierung. Es können außerdem stark komprimierte MRC-PDFs erzeugt werden, die sich insbesondere zur Archivierung von Farbdokumenten eignen.

### Exportieren

Erkennungs- und Umwandlergebnisse können per E-Mail versendet oder an verschiedenen Speicherorten abgelegt werden, wie z. B. in ECM-Systemen, Archiven oder auf SharePoint Servern.\*\*\*

\* Recognition Server Extended Edition bietet im Vergleich zu Professional Edition zusätzlich die Sprachen Thai und Hebräisch. Weitere Sprachen wie Arabisch, Chinesisch, Koreanisch, Japanisch und Vietnamesisch können durch entsprechende Zusatzmodule nachgerüstet werden.

\*\* Unterstützung für die historischen Schriften ist für Recognition Server Extended Edition als Zusatzmodul verfügbar.

\*\*\* Export zu XML steht nur in Recognition Server Extended Edition zur Verfügung.

\*\*\*\* Export zu SharePoint Server steht in Recognition Server Extended Edition zu Verfügung.

# Integration und Verfügbarkeit

## Allgemeine Systemanforderungen:

- **PC** mit Intel® Core™2/2 Quad/ Pentium® / Celeron® / Xeon™, AMD K6/ Turlon™/ Athlon™/ Duron™/ Sempron™ Prozessor mit min. 2 GHz
- **Betriebssystem:** Microsoft® Windows® 7, Windows Server® 2008, Windows Vista®, Windows Server 2003, Windows XP
- **Arbeitsspeicher:** Server Manager: 1 GB RAM; Scanstation: 1 GB RAM; Verarbeitungsstation: 512 MB RAM, sowie zusätzliche 300 MB für jeden Erkennungsprozess; Indexierungsstation: 768 MB RAM
- **Festplattenspeicher:** Server Manager: 20 MB für die Installation, zusätzliche 1 GB für die Programmausführung; Scanstation: 1 GB; Verarbeitungsstation: 600 MB für die Installation, zusätzliche 1 GB für die Programmausführung; Indexierungsstation: 500 MB für die Installation und 1 GB für die Programmausführung
- **Microsoft .NET Framework 2.0** oder höher ist erforderlich, um Dateien in Microsoft SharePoint Server zu speichern.
- **Microsoft Outlook 2000** oder höher für die Verarbeitung von E-Mail-Nachrichten über Microsoft Exchange Server
- **Microsoft IIS 5.1** oder höher (für Web API)
- **Scanner** mit TWAIN-, WIA- oder ISIS-Unterstützung
- **Grafikkarte** und Monitor mit einer Auflösung von 1024 x 768 px

Der für die Programmausführung erforderliche Festplattenspeicherplatz kann je nach Größe, Komplexität und Anzahl der Dokumentenbilder variieren. Die Systemanforderungen sind von den Serverkomponenten oder zusätzlich verwendeten Modulen abhängig. Detaillierte Informationen erhalten Sie von Ihrem ABBYY-Sales Mitarbeiter.

## Eingabeformate:

- BMP, PCX, DCX, GIF, TIFF, WDP
- JPEG, JPEG 2000, JBIG2, PNG
- PDF (up to vers. 1.7), DjVu

## OCR-Sprachen:

Unterstützung von über 190 Sprachen\*\*\*

## Ausgabeformate:

### Editierbare Formate

- RTF, TXT, HTML, CSV, XML, EPUB
- DOC, DOCX (MS Word 2007)
- XLS, XLSX (MS Excel 2007)
- Alto XML

### Durchsuchbare Formate

- PDF (bis PDF 1.7); PDF/A

### Bildformate

- Nur-Bild PDF
- JPEG, JPEG 2000, JBIG2, TIFF

## Integration

ABBYY Recognition Server integriert sich optimal in externe Anwendungen, wie beispielsweise Enterprise-Content-Management und Workflow-Systeme, erweitert vorhandene Anwendungen und automatisiert die Konvertierung und Bereitstellung von Dokumenten. Auch Unternehmen, die Web Services oder andere serviceorientierte Anwendungen implementieren möchten können Recognition Server nutzen.

Recognition Server bietet eine Reihe an Möglichkeiten für die Integration in externe Applikationen. Die Funktionalitäten umfassen:

- XML-Ticket Support\*
- COM-basierte API\*
- Web Service API\*
- Microsoft SharePoint Export\*

## Konnektoren für Enterprise Search\*\*

ABBYY Recognition Server erweitert die Enterprise Search Systeme von Google und Microsoft mit dem Google Search Appliance Konnektor sowie dem IFilter für Microsoft SharePoint Server und Windows Desktop Search. Diese Add-on-Module ermöglichen die Indexierung der Inhalte von Faxen, gescannten Verträgen, Briefen und jedem anderen Dokument in den Formaten TIFF, JPEG oder PDF.

## Lizenzierung

Recognition Server ist in zwei Versionen erhältlich:

### Recognition Server 3.5 Professional Edition

Diese Version bietet Standardfunktionen für Unternehmen, die einen im Hintergrund arbeitenden zentralen OCR-Service benötigen, bei dem keine Integration in andere Anwendungen erforderlich ist.

### Recognition Server 3.5 Extended Edition

Diese Version bietet eine vollständige Integration in externe Anwendungen als Teil einer Web-Service-Architektur. Erweiterte Funktionalitäten der Extended Edition umfassen: zusätzliche OCR-Sprachen (Thailändisch und Hebräisch), XML-Export, Web Service API, COM-basierte API sowie Unterstützung für XML-Tickets und Microsoft SharePoint Export. Erkennung von Fraktur und Erkennung von Arabisch, Chinesisch, Japanisch und Koreanisch stehen als Zusatzmodule zur Verfügung.

### Add-on Module

Beide Editionen können durch das Hinzufügen weiterer Stationen oder mehr CPU-Kerne erweitert werden. Zusätzliche Sprachpakete (Erkennung von Arabisch, Chinesisch, Japanisch, Koreanisch, Vietnamesisch und Hebräisch) und Schnittstellen für Google Search Appliance, IFilter für Microsoft Windows und SharePoint Server Search\*\* erweitern den Funktionsumfang.

### Service und Support

Mit Service und Support erhält der Kunde regelmäßige kostenlose Software-Updates, Upgrades auf die letzte Version sowie auch technischen Support. Diese Leistungen sind beim Software-Kauf obligatorisch, werden separat ausgewiesen und jährlich abgerechnet.

### Testversionen

Wenden Sie sich bitte direkt an uns oder einen unserer Partner, um eine voll funktionsfähige Testversion von ABBYY Recognition Server zu erhalten.

**Hinweis:** Genaue Informationen zu Preisen, verfügbaren Features und Erweiterungsmöglichkeiten der einzelnen Versionen entnehmen Sie bitte unserer Preisliste für ABBYY Recognition Server.

\* Diese Integrationsoptionen sind verfügbar in Recognition Server Extended Edition.

\*\* Anbindungen zu Google Search Appliance (GSA) und iFilter für Microsoft Windows® Search (Desktop und Server) stehen in Recognition Server Extended Edition zur Verfügung. Sie können als Zusatzmodule für den Recognition Server Professional Edition erworben werden.

\*\*\* Recognition Server Extended Edition bietet im Vergleich zu Professional Edition zusätzlich die Sprachen Thai und Hebräisch. Weitere Sprachen wie Arabisch, Chinesisch, Koreanisch, Japanisch und Vietnamesisch können durch entsprechende Zusatzmodule nachgerüstet werden.

# ABBYY®

**ABBYY Europe GmbH**  
Eisenheimerstr. 49  
D-80687 München  
Tel: +49 89 511 159 0  
sales\_eu@abby.com  
[www.ABBYY.com](http://www.ABBYY.com)  
[www.ABBYY.de](http://www.ABBYY.de)

**ABBYY France**  
4, rue Leroux  
94100 Saint-Maur des Fossés  
Frankreich  
Tel: +33 (0)1 41 81 79 71  
sales\_france@abby.com  
[www.france.ABBYY.com](http://www.france.ABBYY.com)

**ABBYY UK Ltd.**  
1210 Parkview, Theale  
Berkshire, RG7 4TY, UK  
Tel: +44 1189 657 640  
sales\_UK@abby.com  
[www.ABBYY.co.uk](http://www.ABBYY.co.uk)

**ABBYY Software Ibérica, S.L.**  
Gran Via de Carles III, 84, 3º  
08028 Barcelona, Spanien  
Tel: +34 93 4965720  
sales\_es@abby.com  
sales\_pt@abby.com