

Service-Beschreibung und Mitwirkungspflichten

Retarus Intelligent Document Processing

Übersicht

Retarus IDP (Intelligent Document Processing) ist eine innovative Dokumentenverarbeitungs-Plattform, die speziell dafür konzipiert wurde, unstrukturierte Geschäftsdokumente (PDF, PNG oder TIFF) effizient in ein strukturiertes XML-Format zu konvertieren.

Die IDP-Plattform nutzt modernste Technologien wie OCR (Optical Character Recognition), trainierte KI-Modelle und einer Kombination aus ML (Machine Learning) und DL (Deep Learning), um relevante Daten aus eingehenden Geschäftsdokumenten wie z.B. Rechnungen oder Bestellungen zu extrahieren und für die automatisierte Weiterverbreitung bereitzustellen.

Retarus IDP funktioniert robust und effizient. Der Service kann große Mengen eingehender unstrukturierter Dokumente erfassen und lässt sich nahtlos mit vorhandenen ERP-Systemen (Enterprise Resource Planning) integrieren.



Unterstützte Formate für eingehende Dokumente

- PDF
- PNG
- TIFF

Unterstützte Dokumenttypen

- Rechnungen
- Bestellungen
- Auftragsbestätigungen
- Lieferscheine

Wesentliche Merkmale

Verarbeitung von Dokumenten

Standardisierter Import

Der Kunde übermittelt unstrukturierte Geschäftsdokumente über verfügbare Standardkanäle (E-Mail, Fax, API-Aufruf) oder mittels eines manuellen Upload-Formulars an die IDP-Plattform. Jedes eingehende Dokument wird auf der Plattform eindeutig identifiziert, mit einer Prozess-ID versehen und mit dem Kunden verknüpft.

Diverse Sicherheitsvorkehrungen schützen vor Missbrauch und sichern die Vertraulichkeit von Daten, sodass unautorisierte Dritte keinen Zugriff auf Kundeninformationen erhalten.

Automatisierte Datenextraktion (AI)

IDP setzt OCR-Technologie ein, um Text, numerische und Datumsinformationen aus hochgeladenen Bildern und PDFs zu extrahieren. Trainierte, dokumententypspezifische KI-Modelle verbessern die Datenerfassung, indem sie die Bedeutung der extrahierten Daten ableiten – auch aus komplexen oder unvollständigen Dokumenten. Jedem intelligent extrahierten Dokumentenelement wird ein Confidence Level zugewiesen, das als Prozentsatz angegeben wird und die Genauigkeit der KI-Interpretation widerspiegelt.

Validierung von Dokumenten

Die extrahierten Daten werden anhand flexibel hinterlegter Geschäftsregeln „presets“ überprüft, um die Datenqualität sicherzustellen. Durch intelligente Interpretation des Kontexts der umgebenden Datenfelder im Dokument lässt sich beispielsweise feststellen, ob der Wert eines Feldes sinnvoll bzw. fragwürdig ist. Validatoren lassen sich für jeden Verarbeitungsablauf individuell konfigurieren „presets“ und entscheiden darüber, ob eine manuelle Prüfung des Dokuments durch einen menschlichen Prüfer erforderlich ist.

Manuelle Dokumentenvalidierung (Human-in-the-Loop)

Bei mehrdeutigen Daten mit einer Interpretationssicherheit unterhalb des festgelegten Schwellwerts oder schlecht lesbaren Dokumenten ermöglicht das Retarus IDP-Portal die manuelle Korrektur und Validierung eingehender Dokumente – ein Prozessschritt, der als „Human-in-the-Loop“ bezeichnet wird. Prüfer können über eine intuitive Benutzeroberfläche die maschinell extrahierten Informationen rasch überprüfen und präzisieren oder fehlerhafte Dokumente ablehnen.

Anreicherung von Stammdaten

Kunden können ihre Stammdaten per API-Aufruf oder SFTP importieren, um die extrahierten Informationen anzureichern und zu verifizieren. Indem die Retarus IDP-Lösung auf Stammdaten referenziert, verbessert diese den Kontext und die Genauigkeit der verarbeiteten Daten. Die Stammdaten können auch bei der manuellen Dokumentenvalidierung sinnvoll unterstützen.

Dokumente „parken“

Das Document Parking dient der automatischen Zwischenspeicherung von Dokumenten, die z. B. aufgrund fehlender Stammdaten die Verarbeitungsvoraussetzungen nicht erfüllen. Ein Cronjob prüft die Voraussetzungen stündlich und gibt das Dokument zur Validierung frei, sobald sie erfüllt sind. Ein Beleg darf allerdings nur eine bestimmte Zeit „geparkt“ werden. Danach wird er automatisch freigegeben und fließt in die Bearbeitung zurück.

Verbesserte Erkennung von Schlüsselwörtern

Kunden können die Bearbeitung von Dokumenten über erkannte Schlüsselwörter automatisieren. Dazu wird eine benutzerdefinierte Liste von Schlüsselwörtern erstellt und jedes Schlüsselwort mit einer vordefinierten Aktion verknüpft. Diese Liste wird als eine spezielle Art von Stammdaten an das System übergeben (siehe oben). Bei Aktivierung der Funktion durchsucht die IDP-Plattform die eingelesenen Dokumente nach diesen Schlüsselwörtern und löst bei Erkennung die jeweils zugeteilte Aktion aus. Wenn mehrere Schlüsselwörter in einem Dokument erscheinen, priorisiert das System die Maßnahmen gemäß der Reihenfolge in der Liste (d.h., die Priorisierung erfolgt "top-down").

Erkannte Schlüsselwörter werden parallel in der Human-in-the-Loop-Benutzeroberfläche sowie innerhalb des Dokuments angezeigt. Zusätzlich wird jedes erkannte Schlüsselwort im XML-Ausgabeergebnis aufgenommen.

Document Review

Die Funktion „External Review“ verbessert die Qualität von Dokumenten durch nahtlose Zusammenarbeit. Benutzer können Dokumente markieren, die einer externen Nachprüfung bedürfen, und detaillierte Notizen zu den festgestellten Problemen hinzufügen. Die Reviewer greifen auf diese Dokumente in einer separaten Übersicht zu und geben dem Kunden direktes Feedback, was einen strukturierten und effizienten Prüfprozess ermöglicht.

Strukturierte XML-Ausgabe

Die verarbeiteten Daten werden in ein strukturiertes XML-Format konvertiert, wodurch sie mit verschiedenen Systemen kompatibel werden. Dies ermöglicht eine nahtlose Integration mit dem Zielsystem. Verarbeitete Dokumente lassen sich via API herunterladen. Abgelehnte Dokumente werden separat gespeichert.

Archivierung

Dieser Background Task archiviert heruntergeladene Dokumente automatisch nach einer vorgegebenen Zeit, um die Systemleistung und die Speicherkapazität zu optimieren. Alle Informationen über die Verarbeitung (Metadaten) bleiben dauerhaft verfügbar; die Dokumente selbst samt zugehöriger Dateien stehen nach der Archivierung nicht mehr direkt zur Ansicht und zum Herunterladen zur Verfügung.

Dashboard und Reporting

Übersicht und Monitoring

Das IDP-Portal bietet einen Einblick in die Systemauslastung für jeden konfigurierten Verarbeitungsablauf „preset“. Kunden können sowohl verarbeitete als auch abgelehnte Dokumente über die Portal-Übersicht prüfen und jedes Dokument und seinen Status auf der Plattform verfolgen.

Reporting

Im Bereich „Reporting“ stehen verschiedene vordefiniertes Reporting zur Verfügung, z. B. die Anzahl der vollständig und automatisch ausgelesenen Felder, die durchschnittliche Verarbeitungsdauer pro Dokument oder pro Seite und weitere wichtige Kennzahlen.

Verwaltung und Konfiguration

Verarbeitungskonfiguration

Voreinstellungen „presets“ definieren die Abfolge von Aktionen, die auf eingehende Dokumente angewendet werden. Sie umfassen Regeln, Zuordnungen und Transformationen. Diese dienen dazu, die Retarus IDP-Lösung an die spezifischen Dokumenttypen und Geschäftsanforderungen des Kunden anzupassen. Kunden können über die intuitive Administrationsoberfläche oder die API mehrere Voreinstellungen je Dokumententyp konfigurieren. Eine Voreinstellung definiert z. B. unterstützte und obligatorische Felder, das minimale Confidence Level pro Feld oder für das Dokument, Datenvervollständigung basierend auf Geschäftsregeln und Stammdaten sowie Validierungsregeln.

Voreinstellungen lassen sich aktualisieren oder deaktivieren.

Benutzerverwaltung

Kunden können ihre Benutzer, Benutzerrollen, Berechtigungen und Zugriffsebenen innerhalb der Retarus IDP-Lösung verwalten. Alternativ kann IDP auch mittels Single Sign-On integriert werden.

Multi-Mandantenfähigkeit und Datensicherheit

Das IDP-System ist als sicheres, mehrmandantenfähiges System konzipiert, das die Datensicherheit gewährleistet.

Mehrsprachiges Portal

Die Benutzeroberfläche der IDP-Anwendung ist aktuell in deutscher (DE) und englischer (EN) Sprache verfügbar.

Flexible Integrationsmöglichkeiten erlauben nahtlose Integration in existierende Prozesse

Alle Hauptfunktionen sind über sichere APIs zugänglich. Alternativ bietet die Retarus IDP-Lösung auch pragmatische Integrationen mit etablierten Branchenstandards (EDI).

Mitwirkungspflichten des Kunden

Übersicht

Dieses Dokument beschreibt die wichtigsten Mitwirkungspflichten und Anforderungen für Kunden, welche die Retarus IDP-Lösung einsetzen. Es gibt einen detaillierten Überblick über Abhängigkeiten, Einflussfaktoren und Verantwortlichkeiten, welche auf Kundenseite erfüllt werden müssen.

Kriterien

1. Stammdaten

Das Erkennungsverfahren hängt von der Qualität der vom Kunden gelieferten Stammdaten ab. Retarus gleicht die extrahierten Daten mit den Stammdaten ab und prüft sie auf Plausibilität. Der Kunde verpflichtet sich, die erforderlichen Stammdaten wie unten beschrieben regelmäßig im CSV-Format via API oder SFTP an das Retarus-IDP-System zu liefern. Die Struktur der Stammdaten (CSV) ist fest vorgegeben und muss der aktuellen Version der [Retarus-Stammdatenrichtlinie](#) entsprechen. Andernfalls ist mit erheblichen Qualitätseinbußen bei der Erkennungsrate zu rechnen. Die Richtlinie wird dem Kunden nach Vertragsabschluss zur Verfügung gestellt. Wir empfehlen die Validierung und Bereitstellung der Stammdaten vor dem Projektstart. Abhängig vom Prozess und dessen Kritikalität empfiehlt Retarus folgende Intervalle für den Datenaustausch:

Dokumententyp: Bestellungen

Mindestens einmal am Tag:

- Produktstammdaten
- Kundenstammdaten
- Materialstammdaten
- Lieferanschriften
- Keywords

Dokumententyp: Auftragsbestätigungen

Mindestens zweimal täglich (z. B. 06:00 Uhr und 12:00 Uhr):

- Lieferantenstammdaten
- Offene Bestellungen (Auftragstabelle)
- Keywords

Dokumententyp: Rechnungen

Mindestens einmal am Tag:

- Lieferantenstammdaten
- Offene Aufträge
- Keywords

Qualität der Stammdaten

- Vergewissern Sie sich, dass alle erforderlichen Datenfelder ausgefüllt sind, keine Werte fehlen und alle Werte korrekt und aktuell sind.
- Verwenden Sie einheitliche und standardisierte Formate für alle Daten, einschließlich konsistenter Verwendung von Datumsformaten, Dezimalzahlen, Währungen, Längen von Textwerten, ...
- Entfernen Sie veraltete, doppelte sowie Dummy- oder Testdaten, um Verarbeitungsfehler zu vermeiden.
- Stellen Sie bitte sicher, dass sich unterschiedliche Arten von Stammdaten nicht überschneiden.

2. Dokumente für das Training

Für ein effizientes Training des KI-Modells, sollten die bereitgestellten Dokumente bestimmte Kriterien erfüllen und zentrale Faktoren berücksichtigen, die Leistung und Qualität beeinflussen. Die Dokumente sollten sich auf häufig auftretende, repräsentative Szenarien konzentrieren, während Ausnahmefälle nur begrenzt berücksichtigt werden. So können präzise und verlässliche Modellergebnisse sichergestellt werden.

Kriterien für vom Kunden zur Verfügung gestellte Musterdokumente

- **Umfassende Abdeckung von Dokumenttypen:**
Kunden müssen repräsentative Beispiele aller Dokumententypen und -formate bereitstellen, die das Modell verarbeiten kann. Der Fokus liegt dabei auf häufig auftretenden Fällen.
- **Berücksichtigung typischer Schwankungen und eines ausreichenden Testvolumens:**
Der Datensatz sollte Variationen von Dokumenten (z. B. unterschiedliche Layouts, Lieferanten/Kunden und Feldplatzierungen) umfassen, welche in der Produktion am häufigsten auftreten.
- **Ausreichende Anzahl an Musterdokumenten:**
Stellen Sie bitte eine angemessene und vielfältige Auswahl an Dokumenten zur Verfügung. Mindestens 500 Dokumente aus den letzten drei Monaten sind erforderlich, um ein effektives und solides Modell-Training zu gewährleisten.
- **Berücksichtigung von Sonderfällen:**
Kunden können auch atypische Dokumente (z. B. Scans mit geringer Qualität oder seltene Vorlagen) einbeziehen. Diese werden auf einem Basisniveau trainiert, jedoch ohne kundenspezifische Anpassungen. Handschriftlicher Text und nicht-lateinische Schriftzeichen werden nur in begrenztem Umfang unterstützt.
- **Qualität und Format der Dokumente:**
 - Bereitstellung von Dokumenten ohne visuelle Verzerrungen (nicht unscharf, nicht verzerrt)
 - Unterstützte Dateiformate sind PDF, PNG, TIFF (mehrsseitig)
 - Maximale Dokumentengröße ist 4 MB / bis zu 50 Mio. Pixel
 - Ein API-Aufruf darf nur ein Dokument enthalten (max. 200 Seiten)
 - Die Mindestbildauflösung beträgt 150 dpi
 - Dokumente sollten im DIN A4-Format vorliegen (sowohl Hoch- als auch Querformat werden unterstützt)

3. Konfiguration

Die Konfiguration spielt eine entscheidende Rolle bei der Optimierung der Automatisierung in der Retarus IDP-Lösung. Obwohl unser Team die Einrichtung übernimmt, basiert diese auf den vom Kunden bereitgestellten Daten und Vorgaben und wird an die Funktionen der IDP-Lösung angepasst. Wir empfehlen, mit den Feldern zu starten, die für den Prozess essenziell sind und bereits in den Dokumenten vorhanden sind.

Nach dem Erhalt eines ersten Dokumenten-Samples, dem Abgleich der Stammdaten und den ersten Verarbeitungsergebnissen führt Retarus eine Analyse durch, um die Produktkonfiguration weiter zu optimieren. Dabei werden dem Kunden die voraussichtlichen Leistungskennzahlen sowie Empfehlungen für eine erste Standardkonfiguration bereitgestellt.