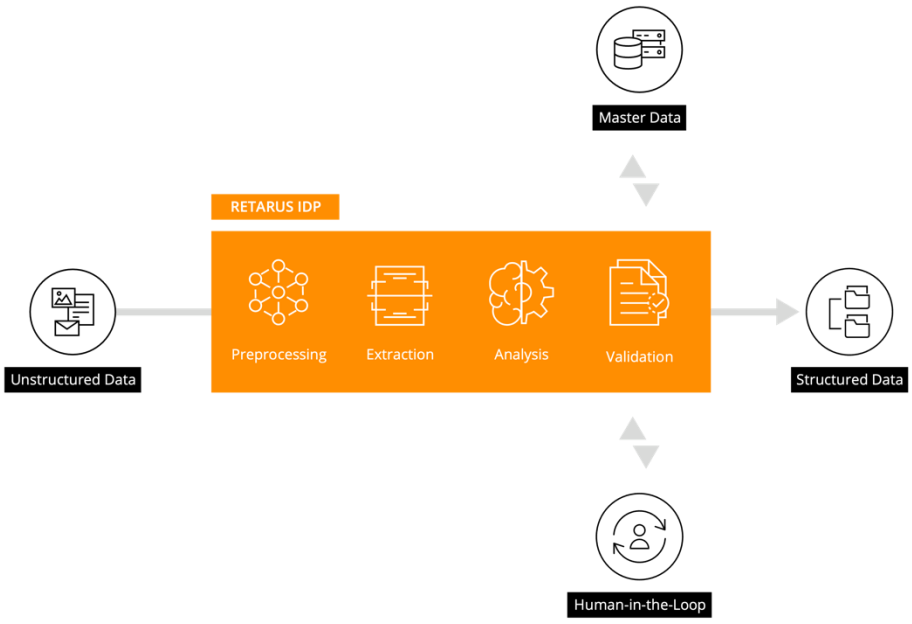
Service Description and Duties of Cooperation
Retarus Intelligent Document Processing

Overview

Retarus IDP (Intelligent Document Processing) is an advanced document processing platform designed to streamline the conversion of unstructured business documents (such as PDF, PNG or TIFF) into structured XML format.

IDP platform leverages cutting-edge technologies, including OCR (Optical Character Recognition), trained AI models, and a combination of ML (Machine Learning) und DL (Deep Learning) to extract relevant data from incoming business documents such are invoices and orders, thus making them available for further automated processing.

IDP offers a robust and efficient solution, whether you are dealing with a high volume of unstructured incoming documents or need to integrate seamlessly with your ERP (Enterprise Resource Planning) system.



Supported incoming document formats

- PDF
- PNG
- TIFF

Supported document types

- Invoice
- Order
- Order Confirmation
- Delivery Note

Key Features

Document Processing

Standardized Import

Customer uploads unstructured business documents to the IDP platform via standard available channels (mail, fax, API call) or via manual upload. Each incoming document is uniquely identified on the platform and associated with a Customer. Various security mechanisms protect against misuse and ensure the confidentiality of data, preventing unauthorized third parties from accessing customer information.

Automated Data Extraction (AI)

IDP employs OCR technology to extract text, numerical and date information from uploaded images and PDFs. Well-trained document type-specific AI models enhance data capture by inferring the meaning of the extracted data, even from complex or imperfect layouts. Each automatically captured document element is assigned a confidence level, expressed as a percentage, indicating the certainty of the captured data accuracy.

Document Validation

Extracted data is run against a set of flexibly configured business rules to ensure data quality, e.g. checking the context of the surrounding data fields in the document to verify if the single field value makes sense, or if the value of the field itself is possible at all. Validators can be customized for each processing flow “preset” to determine whether a document requires human review for verification.

Manual Document Validation (Human-in-the-Loop)

For cases involving ambiguous data, data with capture confidence below the established threshold, or poorly legible documents, our Retarus IDP Portal application facilitates manual correction and validation of incoming documents – the so-called „Human-in-the-loop“ step. Human reviewers use an intuitive user interface to quickly verify and adjust extracted information, ensuring high accuracy, or to reject erroneous documents.

Master Data Enrichment

Customers can import their master data via API call or SFTP to enrich and verify extracted information. By referencing existing master data records, we enhance the context and accuracy of the processed data. Master data can also be used as a tool in manual document validation.

Document Parking

Document Parking automatically sets aside documents that don't meet processing prerequisites, e.g. there is no up-to-date master data available at the document ingestion. A scheduled job checks these prerequisites hourly and releases the document for validation once they are met. A document is allowed to spend just a certain time being parked, after which it is automatically released and put back into processing.

Enhanced Keyword Detection

This feature empowers customers to automate document handling based on keyword detection. Customer can prepare a custom list of keywords, associating each keyword with a pre-defined action. The list is imported into the system as a special type of master data, using the same upload mechanism. When keyword functionality is enabled, the IDP platform automatically scans ingested documents for these keywords and triggers the corresponding action upon detection. In cases where multiple keywords are detected within a single document, the system prioritizes actions based on the order of keywords in the list, with higher-ranking keywords taking precedence.

Detected keywords are displayed in the Human-in-the-Loop interface, showing instance counts and locations within the document. Also, each instance of detected keyword is included in the XML output result.

Document Review

The External Review feature enhances document quality through seamless collaboration. Users can flag documents needing external review and add detailed notes about identified issues. Reviewers access these documents in a separate overview and provide direct feedback to the customer, creating a structured and efficient review process.

Structured XML Output

The processed data is transformed into structured XML format, making it compatible with various systems. This ensures seamless integration with Customer's existing software, such as ERP. Processed documents can be downloaded via API. Rejected documents are stored separately.

Archiving

This background task automatically archives downloaded documents after a set period, optimizing system performance and storage capacity. All information about processing (meta-data) is permanently available, while the documents with associated files eventually become unavailable for viewing and downloading.

Dashboard and Reporting

Overview and Monitoring

Retarus IDP Portal application offers insight into the system workload for each configured processing flow “preset”. Customers can review both processed and rejected documents via the Retarus IDP Portal overview, as well as track each document and its status across the platform.

Reporting

There is a set of predefined reports available on the IDP dashboard, e.g. report on number of full automatically vs. manually processed documents per period, report on ingested documents per preset or overall, report on rejected documents, average duration of the processing, reporting per user, etc.

Administration and Configuration

Processing configuration

Presets define the sequence of actions performed on an input document. They encompass rules, mappings, and transformations designed to tailor IDP to Customer's specific document types and business needs. Customers can easily configure more than one processing preset for each document type by using IDP intuitive administrative interface or by using exposed API. Preset configuration defines supported/mandatory fields, minimum confidence level per field or overall, data completion based on business rules and master data, as well as validation rules.

Presets can be published, updated and deactivated.

User Administration

Customers can manage their users, user roles, permissions, and access levels within the Retarus IDP product. Alternatively IDP does support Single Sign-On integration.

Multitenancy and Data Security

The IDP system is designed as a secure multitenant system ensuring data security.

Multilingual support for Retarus IDP Portal

IDP application is available in several languages: English (EN), German (DE).

Flexible integration options allow seamless integration in your existing processes

Having all the main functionalities exposed through secure APIs, IDP solution offers also pragmatic integration alternatives based on proven industry standards (e.g. EDI).

Duties of Cooperation

Overview

This document outlines the key obligations and requirements for customers adopting the Retarus IDP solution. It provides a detailed overview of dependencies, influencing factors, and responsibilities that must be met on the customer's side to achieve optimal accuracy and ensure the solution's long-term success.

Criteria

1. Master Data

The recognition procedure relies on the quality of the master data provided by the customer. Retarus cross-references the extracted data with the master data and verifies its plausibility. To use this data enrichment the customer is obligated to regularly provide the required master data, as outlined below, in CSV format via API or SFTP integration to Retarus IDP system. The structure of the master data (CSV) is fixed and must be delivered according to the latest version of Retarus Master data guideline. The guideline will be provided to the customer after the contract is signed. We recommend validating and providing the master data prior to the project kick-off. Based on the process and its criticality, Retarus recommends the following intervals for data exchange:

Document type: Customer order

At least once a day:

- Product master data
- Customer master data
- Customer material
- Ship to addresses
- Keywords

Document type: Supplier order confirmations

At least twice a day (e.g., 06:00 and 12:00):

- Supplier master data
- Open orders
- Keywords

Document type: Accounts payable invoices (vendor invoices)

At least once a day:

- Supplier master data
- Open orders
- Keywords

Quality of Master Data

- Ensure that all required data fields are populated, without missing values, correct, and up to date.
- Use uniform and standardized formats across all data, including consistent use of date formats, decimal numbers, currencies, text value lengths, ...
- Remove outdated, redundant, dummy or test data to avoid processing errors.
- Ensure that different types of master data don't overlap.

2. Documents for training

To ensure effective AI model training, the provided documents must meet specific criteria and address key factors that influences efficiency of the model training and quality. These documents should focus on frequently encountered scenarios, with limited consideration for edge cases. By following these guidelines, customers can ensure accurate and reliable model outcomes.

Criteria for Customer-Provided Sample Documents

- **Comprehensive coverage of document types:**
  Customers must provide representative samples of all document types and formats the model will process, focusing on frequently encountered cases.
- **Inclusion of typical variability and sufficient volume of samples:**
  The dataset should include variations of documents (e.g., different layouts, suppliers/customers, and field placements) commonly encountered in production.
- **Sufficient volume of samples:**
  Provide adequate and diverse samples of documents. A minimum of 500 documents from the past 3 months is required to ensure effective and robust model training.
- **Limited representation of edge cases:**
  Customers may include edge case documents (e.g., low-quality scans, rare templates), but these will only be trained to a basic level without custom development. Hand-written text and non-latin letters will have limited support.

- **Document quality and format:**
  - Provide documents without visual distortions (not blurred, not skewed)
  - Supported file formats are PDF, PNG, TIFF (multipage)
  - Maximum document size is 4 MB and up to 50 mil pixels
  - One API request may contain only one document (one document can contain up to 200 pages)
  - Minimum image resolution is 150 DPI
  - documents should be in A4 (both portrait and landscape formats are supported)

3. Configuration

The configuration setup plays a key role in optimizing automation within Retarus IDP solution. While our team manages the setup, it's based on the customer's provided data and guidelines, aligning with IDP solution capabilities. We recommend beginning with a minimum set of fields that are essential and already present in the documents.

After receiving an initial sample of documents, matching master data and expected processing results, Retarus will conduct an analysis to optimize product configuration and provide the customer with expected performance metrics, alongside recommendations for initial preset configuration.