Document Information Extraction

THE BEST RUN **SAP**

Content

1 What Is Document Information Extraction?

Automate your document information extraction processes.

Document Information Extraction helps you to process large amounts of business documents that have content in headers and tables. You can use the extracted information, for example, to automatically process payables, invoices, or payment notes and make sure that invoices and payables match. After you upload a document file to the service, it returns the extraction results from header fields and line items.

> → Tip
>
> - See Supported Document Types and File Formats [page 91].
> - See also Supported Languages and Countries/Regions [page 93].

The service performs the following steps to extract information from the uploaded document file:

1. The document is submitted to optical character recognition.
2. The information from the document is extracted to a JSON file, which the user can query.

For more information, see API Reference [page 109].

You can also use the Document Information Extraction UI to consume the service. See Using the Document Information Extraction UI [page 243] to find out how to subscribe, access, and use the user interface application for the service.

With Document Information Extraction you can:

- Process more documents efficiently with fewer errors and difficulties.
- Increase quality and compliance mechanisms.
- Reduce the time required to process a document.
- Allow the members of your organization to focus on more relevant tasks that are in their field of expertise.

Features

Automate information extraction	Automate the extraction of relevant information from business documents. The Document API takes document files as input and returns header fields and line items as structured data.
Automate data enrichment	Match a business document to enrichment data records based on the information extracted from the document. The Enrichment Data API takes document files as input and returns the ID of the matching enrichment data records.
Benefit from multitenancy support	Use this service in tenant-aware (multitenant) applications. Run them on a shared compute unit that can be used by multiple consumers (tenants).

> ⓘ Note
>
> SAP may continuously improve the above listed core features and their functionalities provided as part of the Document Information Extraction cloud service including automation, transaction processing, and machine learning on behalf of the customer.

> → Tip
>
> Use the data feedback collection feature to allow confirmed documents to be used to improve the Document Information Extraction service.
>
> SAP uses the identity and position of the document-specific fields (see Extracted Header Fields [page 291] and Extracted Line Items [page 300]) as a feedback signal to continuously retrain the machine learning models of the service. With this approach, SAP is able to reduce errors over time when predicting field values from documents.
>
> This is a platform functionality reused by other applications. SAP reserves the right to reject documents submitted for retraining.
>
> For more information, see Create Configuration [page 122], Confirm Document [page 165] and Data Protection and Privacy [page 302].

Environment

This service is available in the following environments:

- Cloud Foundry environment
- Kyma environment

Multitenancy Support

This service supports multitenancy. It can be used in tenant-aware applications.

For information on multitenancy support, see Run the Service in a Multitenant Application [page 107].

Prerequisites

See Initial Setup [page 103].

Technical Constraints

For information on technical limits, see Technical Constraints [page 288].

Regional Availability

Get an overview on the availability of Document Information Extraction according to region, infrastructure provider, and release status in the *Pricing* tab of the SAP Discovery Center .

Trial Scope

Document Information Extraction is available for trial use. A trial account lets you try out SAP Business Technology Platform (SAP BTP) for free and is open to everyone. Trial accounts are intended for personal exploration, and not for productive use or team development. They allow restricted use of the platform resources and services.

To activate your trial account, go to Welcome to SAP BTP Trial.

> ⓘ Note
>
> See also the following information: Trial Accounts and Free Tier.

In the Cloud Foundry environment, you get a free trial account for Document Information Extraction with the following constraints: Free Tier Option and Trial Account Technical Constraints [page 289].

2   What's New for Document Information Extraction

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Extracted Header Fields - *quantity* Deprecation	We plan to update the list of header fields that you can extract from `purchaseOrder` documents. The *quantity* header field will be deprecated and will no longer be available for extraction from September 2024. See Extracted Header Fields [page 291].	Info only	Deprecated	Announcement	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-07-08	2024-09-20
Document Information Extraction	Cloud Foundry	Improved Performance of Template-Based Extraction for Larger Schemas	We've improved the performance of template-based extraction. If your schemas include many templates, extraction is now faster than it was in the past.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-07-29	2024-07-29
Document Information Extraction	Cloud Foundry	Extracted Header Fields	You can now extract the *invoiceReferenceNumber* header field from `invoice` documents. The *invoiceReferenceNumber* header field is only relevant for e-invoices from India and is extracted from a barcode in the documents. See Extracted Header Fields [page 291].	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-07-29	2024-07-29

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Purchase Orders - Unit of Measure Mapping	The mapping of unit of measure specifications to standard codes in `purchaseOrder` documents has been improved.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-07-29	2024-07-29
Document Information Extraction	Cloud Foundry	Schema Description in the Prompt for Document Processing with LLMs	You can now use the schema description value to get better results for extractions with generative AI. The schema description value is now used in the prompt for document processing with large language models (LLMs). ⚠ **Restriction** This option is available for schemas with the service plan **Document Information Extraction, premium edition** (`premium_edition`) **only**. See Service Plans [page 84] and Metering and Pricing [page 86]. See Create Schema [page 193] (API documentation), Create Schema [page 257] (UI documentation), and Extraction Using Generative AI: Best Practices [page 286].	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-07-15	2024-07-15

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Schema Field Categories	You can now create categories, which you can use to group data fields in schemas. If your schemas include many data fields, using categories helps avoid long lists of field names, making it easier to find specific fields. To use the *Schema Field Categories* feature, you must first enable it under *UI Settings*. See Schema Field Categories [page 263] and Enable Features [page 247].	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-07-15	2024-07-15
Document Information Extraction	Cloud Foundry	New Panels for Creating, Viewing, Editing, and Deleting Data Fields in Schemas	When you create data fields for schemas, you now make your entries on a panel. This panel replaces the dialog used to date. The *Field Name* entries under *Header Fields* and *Line Item Fields* are now links. To view the details of a field, simply click on the link. A panel appears, where you can also edit or delete the field by choosing the respective button. See Add Data Fields [page 259].	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-07-15	2024-07-15
Document Information Extraction	Cloud Foundry	Improved Extraction of Line Items Directly above Document Footers	We've fixed an issue with the extraction of line items. If the line spacing between the final line item and the document footer is the same as the line spacing in the line items, Document Information Extraction no longer extracts the footer text as part of the final line item.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-07-08	2024-07-08

Technical Component	Environment	Title	Description	Action	Life-cycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Schema Description and Schema Field Description	We've increased the maximum length of the following strings to 500 characters: • `schemaDescription` • schema field `description` See Create Schema [page 193] and Add Fields to Schema Version [page 208].	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-07-08	2024-07-08
Document Information Extraction	Cloud Foundry	New Improved ML Models for Invoice, Purchase Order, and Payment Advice.	We've improved the machine learning models for the document types invoice, purchase order, and payment advice.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-07-01	2024-07-01
Document Information Extraction	Cloud Foundry	New SAPUI5 Version for the Document Information Extraction UI	The Document Information Extraction UI now uses SAPUI5 Version 1.122. Notable changes include resizable panels in the document viewer.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-06-24	2024-06-24
Document Information Extraction	Cloud Foundry	Line-Item Matching in Templates	We've improved line-item matching in templates by lowering a threshold for the overlap between bounding boxes and content.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-06-24	2024-06-24

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Separator for 1000 and Decimal Separator in Templates	We've improved the template logic for determining whether a comma is a 1000 separator or a decimal separator in numerical values.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-06-24	2024-06-24
Document Information Extraction	Cloud Foundry	Extraction Using Template	We've fixed an issue with extracting information using templates. Adding multiple documents to a template and annotating different fields across these documents no longer prevents Document Information Extraction from extracting information using that template.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-06-24	2024-06-24
Document Information Extraction	Cloud Foundry	New Improved ML Models for Invoice, Purchase Order, and Payment Advice.	We've improved the machine learning models for the document types invoice, purchase order, and payment advice.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-06-24	2024-06-24

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Identifier API	The Identifier API (Deprecated) [page 117] and its endpoints are now deprecated and scheduled for decommissioning in June 2025. Following that date, the Identifier API and the following endpoints will no longer be available: • Create Identifier (Deprecated) [page 118] • Get Identifier (Deprecated) [page 120] • Delete Identifier (Deprecated) [page 121] Please use the Schema API [page 193] to process `paymentAdvice` documents in Excel format.	Recommended	Deprecated	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-06-24	2024-06-24
Document Information Extraction	Cloud Foundry	Tooltips with Enrichment Metadata on the Document Information Extraction UI	If you use the enrichment feature when adding documents to the Document Information Extraction service, you can now view enrichment metadata on the Document Information Extraction UI. This metadata shows you the enrichment method, allowing you to debug any issues with your master data more easily. To display the metadata, simply hover your cursor over the relevant entry in the *Enrichment* section under *Extraction Results*.	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-06-24	2024-06-24

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Simplified Document Upload on the Document Information Extraction UI	When you add documents to the Document Information Extraction UI, it's now mandatory to include a schema. As a result, you no longer have to review fields for extraction before confirming the document upload – making for quicker and easier uploads.	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-06-24	2024-06-24
Document Information Extraction	Cloud Foundry	Upload Document - Options Payload - New Option: `schemaName`	When uploading a document to the Document Information Extraction service, you can now add `schemaName` to the *Options Payload*. Previously, only the `schemaId` option was available. You can also use `schemaName` (instead of schemaId) to add one of the following preconfigured SAP schemas to the *Options Payload*: • `SAP_OCROnly_schema` • `SAP_invoice_schema` • `SAP_purchaseOrder_schema` • `SAP_paymentAdvice_schema` See Upload Document [page 135].	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-06-24	2024-06-24

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Template API	When creating metadata for a template, you can now add to the payload the options `extraction` and `isLineItemField` to provide extraction model information for a field. See Create Template Metadata [page 232] and Get Template Metadata [page 233].	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-06-24	2024-06-24
Document Information Extraction	Cloud Foundry	Enrichment Data API	The Enrichment Data API [page 174] endpoint Delete Enrichment Data (Synchronous) - Deprecated [page 189] is now deprecated and scheduled for decommissioning in November 2024. Following that date, the endpoint will no longer be available. Please use the endpoint Delete Enrichment Data (Asynchronous) [page 191] to delete data records.	Recommended	Deprecated	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-05-13	2024-05-13
Document Information Extraction	Cloud Foundry	Better Models for the Extraction of Standard Document Types	The machine learning models for the extraction of `invoice`, `paymentAdvice`, and `purchaseOrder` documents have been improved.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-05-13	2024-05-13
Document Information Extraction	Cloud Foundry	Better Extraction of `rawValue` for Standard Document Types and Fields	The extraction of the `rawValue` response field has been improved for the standard document types and fields. See Get Result [page 146].	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-05-13	2024-05-13

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Extracted Line Items	You can now extract purchase order numbers that are available on line-item field level from `invoice` documents. See Extracted Line Items [page 300].	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-03-11	2024-03-11
Document Information Extraction	Cloud Foundry	Configuration API	You can now use the `client` scope configuration for the `dataFeedbackCollection` configuration key. See Configuration Keys [page 124].	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-03-11	2024-03-11
Document Information Extraction	Cloud Foundry	Post Catalog	You can now filter documents based on `schemaId`. See Post Catalog [page 142].	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-03-11	2024-03-11
Document Information Extraction	Cloud Foundry	New Invoice Supported Language - Japanese	The Document Information Extraction service now supports the Japanese language for `invoice` documents. See Invoice: Languages and Countries/Regions [page 94].	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-03-11	2024-03-11

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Better Models for the Extraction of Standard Document Types	The machine learning models for the extraction of `invoice`, `paymentAdvice`, and `purchaseOrder` documents have been improved.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2024-03-11	2024-03-11
Document Information Extraction	Cloud Foundry	Better Extraction of Line Items from Multi-page Documents with Template	The template algorithm has been enhanced. Document Information Extraction now delivers better results when extracting line items from multi-page documents with a table header that appears only on the first page.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2024-02-20	2024-02-20

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Combine Different Setup Types When Adding Data Fields to Schemas	You can now combine header fields with different setup types in the same schema. You can add header fields with the following setup types to a schema created for a **standard document type**: • *auto* (with and without a default extractor) • *manual* You can add header fields with the following setup types to a schema created for a **custom document type**: • *auto* (without a default extractor) • *manual* ⚠ **Restriction** The setup type *auto* is available without default extractor (extraction using generative AI) for schemas with the service plan **Document Information Extraction, premium edition** (`premium_edition`) **only**. See Service Plans [page 84] and Metering and Pricing [page 86]. See also Add Fields to Schema Version [page 208], Add Data Fields [page 259], and Setup Types [page 261].	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-02-20	2024-02-20

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Invoices - Conversion of Country Specific Unit of Measure Values to ISO Format	The conversion of country specific unit of measure values to ISO format for `invoice` documents has been improved.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-02-20	2024-02-20
Document Information Extraction	Cloud Foundry	Support for `businessCard` Documents in AWS region Australia (Sydney)	The `businessCard` documents are now supported in the AWS region Australia (Sydney). See Supported Document Types and File Formats [page 91].	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-02-20	2024-02-20
Document Information Extraction	Cloud Foundry	Download Troubleshooting Data for Documents	You can now download data about documents added to the Document Information Extraction UI for use in troubleshooting any issues. See Download Troubleshooting Data [page 252].	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-02-05	2024-02-05
Document Information Extraction	Cloud Foundry	Model Used for Extraction	The Document API now includes information about the model used for extraction. As a result, you can see whether Document Information Extraction used a template or AI to extract information from a particular field. See Get Result [page 146].	Info only	General Availability	New	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-02-05	2024-02-05

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	New Location for Schema Configuration Feature on UI	You now call up the Schema Configuration feature of the Document Information Extraction UI directly from the navigation bar on the left of the screen. See Create Schema [page 257].	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-02-05	2024-02-05
Document Information Extraction	Cloud Foundry	Extraction of Descriptions from Columns	We've fixed an issue with extracting description values from columns. Document Information Extraction now extracts the complete content of large column cells containing descriptions of numbers or quantities, for example.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-02-05	2024-02-05
Document Information Extraction	Cloud Foundry	Extraction of Line Items	We've fixed an issue with extracting line items. If the template returns the extraction result invalid, but the AI returns the extraction result valid for the same line item, the final result is now valid when Document Information Extraction merges the two results.	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-02-05	2024-02-05
Document Information Extraction	Cloud Foundry	Get Templates Endpoint	The `limit` parameter of the Get Templates endpoint is now independent of the `order` parameter. To apply the `limit` parameter, you no longer need to specify a value for `order`. See Get Template [page 221].	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-02-05	2024-02-05

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Display Description for Fields in Extraction Results	You can now display the description text for fields in the *Extraction Results* pane on the Document Information Extraction UI. To view the description, open the *Extraction Results* pane and hover over the name of a header field or line item. A tooltip appears, which includes the description text. See View and Edit Extraction Results [page 252].	Info only	General Availability	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-02-05	2024-02-05
Document Information Extraction	Cloud Foundry	Extracted Line Items - *materialNumber* and *senderMaterialNumber* Deprecation in `SAP_purchaseOrder_schema`	The line items *materialNumber* and *senderMaterialNumber* were replaced by *supplierMaterialNumber* and *customerMaterialNumber* respectively in the list of fields that you can extract from `purchaseOrder` documents when using the `SAP_purchaseOrder_schema`. The legacy line items *materialNumber* and *senderMaterialNumber* are now deprecated and no longer available for `purchaseOrder` documents. See Extracted Line Items [page 300].	Info only	Deprecated	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-02-05	2024-02-05

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Extracted Line Items - *currencyCode* Deprecation	We updated the list of line items that you can extract from `purchaseOrder` documents. The *currencyCode* line item is now deprecated and no longer available for extraction. See Extracted Line Items [page 300].	Info only	Deprecated	Changed	Intelligent Technologies	Not applicable	SAP Business Technology Platform	2024-02-05	2024-02-05

2.1 2023 What's New for Document Information Extraction (Archive)

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of

Document Information Extraction	Cloud Foundry	Prefilled Setup Types for Schema Fields	When you add data fields to schemas, the service now prefills the *Setup Type* field with default values.
			
			Depending on whether you use Document Information Extraction, premium edition or base edition, the default values are as follows:
			
			• Premium edition
			  • Schemas for standard and custom document types: *auto*
			• Base edition
			  • Schemas for standard document types: *auto*
			  • Schemas for custom document types: *manual*
			
			See Setup Types [page 261].
				Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-12-11	2023-12-11

Document Information Extraction	Cloud Foundry	Support for X.509 Authentication	The Document Information Extraction APIs now support X.509 authentication. See Enable X.509 Authentication [page 105].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-12-11	2023-12-11
Document Information Extraction	Cloud Foundry	Auditing and Logging Information	New client related events have been created. See Auditing and Logging Information [page 305].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-12-11	2023-12-11

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Template API	From now, you can't download documents that are part of the template export package but haven't been malware-scanned during upload. You can download malware-scanned documents only. See Export Template [page 231].	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-12-11	2023-12-11
Document Information Extraction	Cloud Foundry	Document Information Extraction UI	There have been several security improvements on the Document Information Extraction UI. See Using the Document Information Extraction UI [page 243].	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-12-11	2023-12-11

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	New Service Plan: **Document Information Extraction, premium edition** (`premium_edition`)	The service plan **Document Information Extraction, premium edition** (`premium_edition`) is now generally available. The `premium_edition` service plan allows you to use **generative AI** to automate use cases for business document processing with **large language models (LLMs)**. Use **generative AI** to process business documents in more than 40 languages, and implement new business document use cases with shorter time to value. You can also use an SAP BTP trial account to try out the document information extraction using **generative AI**. Follow the tutorial: Use Trial to Extract Information from Custom Documents with Generative AI and Document Information Extraction . See Service Plans [page 84] and Metering and Pricing [page 86]. See also Extraction Using Generative AI: Languages [page 101], Add Fields to Schema Version [page 208], Setup Types [page 261], and Extraction Using Generative AI: Best Practices [page 286].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-12-06	2023-12-06

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Template API	The Template API [page 219] is now generally available. You can now use the Template API endpoints to create, reuse, edit, and delete templates based on schemas and document types.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-11-27	2023-11-27
Document Information Extraction	Cloud Foundry	Machine Translation available for the Document Information Extraction SAP Help Portal Documentation	For your convenience, machine translation from the original and official English language is now available for the Document Information Extraction documentation on SAP Help Portal in the following languages: • Chinese Simplified • French • German • Italian • Japanese • Korean • Portuguese • Spanish	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-11-27	2023-11-27
Document Information Extraction	Cloud Foundry	Configuration API and Notifications	In addition to the already available `instance` and `tenant` scopes, you can now also use the `activateDocumentNotifications` configuration key on `client` scope level to enable the Notifications [page 236] functionality and get notifications about the status of your processed documents. See Configuration Keys [page 124].	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2024-01-08	2023-11-27

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Better Model for the Extraction of `invoice` Documents	The machine learning model for the extraction of `invoice` documents has been improved. The improvements include better extraction results for currency, country and date fields. Additionally, the service now supports the following countries/regions for Invoice: Languages and Countries/Regions [page 94] documents (and their corresponding languages): • Hungary (Hungarian) • Romania (Romanian) • Türkiye (Turkish)	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-11-27	2023-11-27
Document Information Extraction	Cloud Foundry	Better Model for the Extraction of `paymentAdvice` Documents	The machine learning model for the extraction of `paymentAdvice` documents has been improved. The improvements include better extraction results for currency and country fields.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-11-27	2023-11-27
Document Information Extraction	Cloud Foundry	Better Model for the Extraction of `purchaseOrder` Documents	The machine learning model for the extraction of `purchaseOrder` documents has been improved. The improvements include better extraction results for currency, country and date fields. Additionally, the service now supports the extraction of quantities with multipliers, for example, "2x5".	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-11-27	2023-11-27

Technical Component	Environment	Title	Description	Action	Life-cycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Enrichment Data API	The `orderby` parameter was replaced by `order` in December 2022. The legacy `orderby` parameter is now deprecated and no longer available. See List Data-Persistence Jobs [page 182].	Info only	Deprecated	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-11-27	2023-11-27
Document Information Extraction	Cloud Foundry	New Generative AI Tutorial	The tutorial Use Trial to Extract Information from Custom Documents with Generative AI and Document Information Extraction is now available. Learn how to use Document Information Extraction with generative AI to automate the extraction of information from custom document types using large language models (LLMs).	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-11-10	2023-11-10
Document Information Extraction	Cloud Foundry	Data Feedback Collection for Model Improvement	You can now use the feedback collection feature in the Document Information Extraction UI to consent to the use of confirmed documents to retrain the service's machine learning models. See Confirm Documents [page 255].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-11-05	2023-11-05
Document Information Extraction	Cloud Foundry	Document Information Extraction UI	The look and feel of the Document Information Extraction UI has been updated to provide the latest SAP Fiori user experience.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-11-29	2023-11-05

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Edit Template	You can now edit templates. In addition to changing the name and description, you can choose a different schema for the template. Changing the schema makes a new set of extraction fields available for the template. If you've already edited extraction results for sample documents associated with your template, these edits are preserved following the change of schema if the relevant fields appear in both the old and the new schema. See Edit Template [page 269].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-11-29	2023-10-23
Document Information Extraction	Cloud Foundry	Field Label	In *Schema Configuration*, you can now optionally enter a field label in the *Add Data Field* dialog. These labels enable you to give user-friendly names to some or all of the header fields and line item fields that you add to schemas. Field labels that you define in this way are displayed instead of the technical field names under *Extraction Results* in the *Document* feature of the Document Information Extraction UI. See Add Data Fields [page 259].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-10-23	2023-10-23

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Better Model for the Extraction of `invoice` Documents	The machine learning model for the extraction of `invoice` documents has been improved. The improvements include better extraction results for date fields.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-10-23	2023-10-23
Document Information Extraction	Cloud Foundry	Better Model for the Extraction of `paymentAdvice` Documents	The machine learning model for the extraction of `paymentAdvice` documents has been improved. The improvements include better extraction results for date fields, and amount fields in line items.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-10-23	2023-10-23
Document Information Extraction	Cloud Foundry	Built-In Support	You can now use the integrated Built-In Support tool to quickly find answers to your support-related questions. Built-In Support is an embedded digital assistant that allows you to search for support-related information without leaving the UI. If you have an s-user ID and the associated authorizations, Built-In Support also allows you to report issues, review cases, and chat with an expert or a chatbot. See Built-In Support [page 249].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-10-09	2023-10-09

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Configuration API	The `enrichmentConfidenceThreshold` configuration key is now available. You can now adjust the similarity confidence threshold for the enrichment. See Create Configuration [page 122], Configuration Keys [page 124], and Enrichment Data API [page 174].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-10-09	2023-10-09
Document Information Extraction	Cloud Foundry	New Autosave Feature for Editing Extraction Results	You can now have the Document Information Extraction UI save your edits to extraction results. When you choose *Autosave* on the *Edit Extraction Results* pane in the *Documents* feature, the service saves your work automatically at 10-second intervals. See View and Edit Extraction Results [page 252].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-10-09	2023-10-09
Document Information Extraction	Cloud Foundry	New Schema Field Setup Types	The setup types *auto* and *manual* are now available when you add data fields to new schemas. See Add Fields to Schema Version [page 208] and Add Data Fields [page 259].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-10-09	2023-10-09
Document Information Extraction	Cloud Foundry	Technical Constraints	You can now associate a maximum of 5 documents with a template. See Technical Constraints [page 288], Free Tier Option and Trial Account Technical Constraints [page 289] and Add Documents and Activate/Deactivate Template [page 267].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-10-09	2023-10-09

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Associated Confirmed Documents with Templates	You can now associate documents that have the status "CONFIRMED" with templates. If you edit the extraction results for a document and then confirm the document, you can use the *Add to Document* feature to associate the document with a template. See Add Documents and Activate/Deactivate Template [page 267].	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-10-09	2023-10-09
Document Information Extraction	Cloud Foundry	Better Model for the Extraction of `purchaseOrder` Documents	The machine learning model for the extraction of `purchaseOrder` documents has been improved. The improvements include better extraction results for date fields and better formatting of amounts.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-10-09	2023-10-09
Document Information Extraction	Cloud Foundry	Role Collections	The role collection `Document_Information_Extraction_UI_Admin_User` has been deprecated. To create or delete schemas and templates, use the role collection `Document_Information_Extraction_UI_Templates_Admin`. See Role Collections [page 245].	Info only	Deprecated	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-10-09	2023-10-09

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Use Generative AI to Extract Information from Standard and Custom Document Types	You now have the option of using generative AI to extract information from standard and custom document types. To use generative AI, select the setup type *auto* without a default extractor when adding data fields to a schema for a standard or custom document type. ⚠ **Restriction** This option is currently available in SAP BTP trial accounts only. If you don't want to use generative AI with standard or custom document types, select the setup type *manual* when adding fields to schemas. With standard document types, you can also avoid using generative AI by selecting *auto* with a suitable default extractor. See Add Fields to Schema Version [page 208] and Setup Types [page 261].	Info only	Restricted Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-11-10	2023-10-05

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Schema API - Add Schema Fields	You can now optionally use the `label` property to enter field labels. These labels enable you to give user-friendly names to some or all of the `headerFields` and `lineItemFields` that you include in the Add Fields to Schema Version [page 208] `payload`. Field labels that you define in this way are displayed instead of the technical field names under *Extraction Results* in the *Document* feature of the Document Information Extraction UI.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-09-04	2023-09-04
Document Information Extraction	Cloud Foundry	Free Tier Option and Trial Account Technical Constraints	Free tier and trial account users can now: • Upload up to 50 document pages per tenant in a rolling period of 30 days. • Create up to 1000 schemas per client. See Free Tier Option and Trial Account Technical Constraints [page 289].	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-09-04	2023-09-04
Document Information Extraction	Cloud Foundry	Extraction Results Saved Automatically when Documents Associated with Templates	You no longer need to save extraction results manually before associating documents with templates. The Document Information Extraction UI now saves these results automatically.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-08-18	2023-08-18

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Schema API	The Schema API [page 193] is now generally available. You can now use the Schema API endpoints to create, list, update, and delete schemas and schema versions.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-08-18	2023-08-18
Document Information Extraction	Cloud Foundry	Technical Constraints	The maximum total number of header fields and line items you can add per schema is now 500. See Technical Constraints [page 288].	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-08-18	2023-08-18
Document Information Extraction	Cloud Foundry	Better Model for the Extraction of `invoice` Documents	The machine learning model for the extraction of `invoice` documents has been improved. The improvements include better extraction results for bank account numbers, amounts with non-standard formats and numerical dates with whitespaces.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-08-18	2023-08-18
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-08-18	2023-08-18

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Delete Bounding Boxes	When editing extraction results with the Document Information Extraction UI, you can now delete bounding boxes together with their coordinates. See View and Edit Extraction Results [page 252].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-07-26	2023-07-26
Document Information Extraction	Cloud Foundry	Display and Edit Bounding Boxes	When editing extraction results with the Document Information Extraction UI, you can now open the *Assign Field* dialog for bounding boxes by choosing the relevant tooltip in the page preview pane.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform		2023-07-26
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-07-26	2023-07-26
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-07-17	2023-07-17

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Technical Constraints	The maximum number of templates you can create has been increased from 1000 templates per tenant to 1000 templates per schema. See Technical Constraints [page 288].	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-06-30	2023-06-30
Document Information Extraction	Cloud Foundry	Support for Country Code Conversion in Template	The Template [page 266] feature supports now country code conversion.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-06-22	2023-06-22
Document Information Extraction	Cloud Foundry	New Data Type *country/region* for Schema Fields	The new data type *country/region* is now available for schema fields. See Add Data Fields [page 259].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-06-22	2023-06-22
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-06-13	2023-06-13

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Issues with Units of Measure in `purchaseOrder` Documents Corrected	Some issues with codes for units of measure in `purchaseOrder` documents have now been resolved.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-06-13	2023-06-13
Document Information Extraction	Cloud Foundry	Support for Bounding Boxes around Parts of Fields	When you edit extraction results, you can now draw bounding boxes around parts of header field entries, instead of around the entire entry. As a result, you can eliminate unwanted or irrelevant elements, such as punctuation, from strings and ensure that they include only the values that you need. See View and Edit Extraction Results [page 252].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-06-30	2023-06-13
Document Information Extraction	Cloud Foundry	Better Model for the Extraction of `paymentAdvice` Documents	The machine learning model for the extraction of `paymentAdvice` documents has been improved.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-06-13	2023-06-13
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-05-23	2023-05-23

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	*Setup Type* field on *Add Data Field* dialog for schemas	The *Add Data Field* for schema configuration now includes a new field: *Setup Type*. See the updated procedure in Add Data Fields [page 259].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-05-08	2023-05-08
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-05-08	2023-05-08
Document Information Extraction	Cloud Foundry	Response Field clientId in Get Result Endpoint	The Document API endpoint Get Result [page 146] includes a new response field: clientId. You can now identify the client that submitted the extraction request using the Upload Document [page 135] endpoint.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-05-08	2023-05-08
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-04-20	2023-04-20

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Fixed Values in Template Extraction Fields	You can now include fixed values for selected extraction fields in a template. If you intend to use a template with documents from only one supplier, for example, you can define the supplier's name as the fixed value for the *senderName* field. See Add Template [page 266].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-04-04	2023-04-04
Document Information Extraction	Cloud Foundry	Scene Text Recognition Schema	You can now extract text from images using the OCR engine for scene text recognition. When you create a schema with the document type *Custom*, you can choose between two types of OCR engine (*Document* or *Scene Text*), depending on whether the text you wish to extract is in an image or not. See Schema Configuration [page 256] and Create Schema [page 257].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-04-04	2023-04-04
Document Information Extraction	Cloud Foundry	Filtering, ordering, and pagination	The new Document API endpoint Post Catalog [page 142] is now available. You can use the following catalog options to get a list with all document processing jobs in a JSON file: • Filtering • Ordering • Pagination The Document Information Extraction UI also supports document filtering, ordering, and pagination.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-04-04	2023-04-04

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Configuration API and Notifications	The `activateDocumentNotifications` configuration key is now available. You can now enable the Notifications [page 236] functionality to get notifications about the status of your processed documents. See Create Configuration [page 122].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-04-04	2023-04-04
Document Information Extraction	Cloud Foundry	New Procedure for Associating Documents with Templates	There's now a new procedure for adding documents to templates on the Document Information Extraction UI. In the past, you selected these documents when creating the template or added them later using the *Template* feature. Now, you select documents using the new *Add to Template* function in the *Document* feature. See Add Documents and Activate/Deactivate Template [page 267] and View and Edit Extraction Results [page 252].	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-04-04	2023-04-04
Document Information Extraction	Cloud Foundry	Better Model for the Extraction of `invoice` Documents	The machine learning model for the extraction of `invoice` documents has been improved.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-04-04	2023-04-04

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-04-04	2023-04-04
Document Information Extraction	Cloud Foundry	Get Templates Endpoint	The new Document API endpoint Get Templates Associated with Document [page 173] is now available. You can get all the templates associated with the specified document ID.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-03-14	2023-03-14
Document Information Extraction	Cloud Foundry	New Template Feature Supported Language - Greek	The Document Information Extraction UI Template [page 266] feature supports now the Greek language. See Extraction Using Template: Languages [page 99].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-03-14	2023-03-14
Document Information Extraction	Cloud Foundry	Better Model for the Extraction of purchase Order Documents	The machine learning model for the extraction of `purchaseOrder` documents has been improved.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-03-14	2023-03-14

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-03-14	2023-03-14
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements. The performance of the Template [page 266] feature has been improved.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-03-01	2023-03-01
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-02-17	2023-02-17
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code, security, and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2023-02-06	2023-02-06

Technical Component	Environment	Title	Description	Action	Life-cycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Barcode Header Field Symbology	You can now see in the response from Get Result [page 146], in the `symbology` response field, the type of the extracted barcode header fields.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2023-01-30	2023-01-30

2.2 2022 What's New for Document Information Extraction (Archive)

Technical Component	Environment	Title	Description	Action	Life-cycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Configuration API	The `coordinateFormat` configuration key is now available. You can now choose the format of the bounding box coordinates in the extraction results. See Create Configuration [page 122].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-12-19	2022-12-19

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Enrichment Data API	The `orderby` parameter has been replaced by `order`. ⓘ Note The legacy `orderby` parameter will still be supported for a limited amount of time. Please start using the new parameter (`order`) as soon as possible. See List Data-Persistence Jobs [page 182].	Recommended	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-12-19	2022-12-19
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-12-19	2022-12-19
Document Information Extraction	Cloud Foundry	Document Information Extraction UI	The Document Information Extraction UI and associated in-app help are now available in the following new languages: • Chinese Simplified • Chinese Traditional • French • Italian • Japanese • Korean • Portuguese • Russian • Spanish See Set Screen Language [page 248].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-12-07	2022-12-07

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Enrichment Data Method	You can now see in the response from Get Result [page 146], in the `method` response field, the match strategy used for each matched enrichment data record.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-12-07	2022-12-07
Document Information Extraction	Cloud Foundry	Change Service Instance by Name	You can now change instances on the Document Information Extraction UI by entering the service instance name. See Subscribing to the Document Information Extraction UI [page 243].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-12-07	2022-12-07
Document Information Extraction	Cloud Foundry	Better Model for the Extraction of `purchaseOrder` Documents	The machine learning model for the extraction of `purchaseOrder` documents has been improved.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-12-07	2022-12-07
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-12-07	2022-12-07

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Document Information Extraction UI	The Document Information Extraction UI and associated in-app help are now available in German. See Set Screen Language [page 248].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-11-15	2022-11-15
Document Information Extraction	Cloud Foundry	SAP Schemas	The preconfigured SAP schema `SAP_OCROnly_schema` is now available for custom documents and OCR (Optical Character Recognition) output only. See Upload Document [page 135], Get Result [page 146], and Add Document [page 251].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-11-09	2022-11-09
Document Information Extraction	Cloud Foundry	Configuration API	You can now use the `client` scope configuration for the `documentRetentionTimeDays` configuration key. You can now use the optional parameters `clientId` and `tenantId` to create, get and delete configurations. See Create Configuration [page 122].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-11-09	2022-11-09
Document Information Extraction	Cloud Foundry	Free Service Plan	The Template [page 266] feature is now also available to Free service plan users.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-11-09	2022-11-09

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-11-09	2022-11-09
Document Information Extraction	Cloud Foundry	Role Collections	The role collection `Document_Information_Extraction_UI_Templates_Admin` now includes permissions for reading and writing documents. See Role Collections [page 245].	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-10-04	2022-10-04
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-10-04	2022-10-04
Document Information Extraction	Cloud Foundry	Enrichment Data API	The following `paymentAdvice` fields now support enrichment: • *taxId* • *senderAddress* • *senderName* See Extracted Header Fields [page 291].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-09-13	2022-09-13

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-09-13	2022-09-13
Document Information Extraction	Cloud Foundry	Role Collections	The role collection `Document_Information_Extraction_UI_Document_Viewer` is now available. This new collection allows users to read documents in the UI application. See Role Collections [page 245].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-08-30	2022-08-30
Document Information Extraction	Cloud Foundry	Client Segregation	You can now restrict user access to specified clients. See Create Configuration [page 122] and Add Document [page 251].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-08-30	2022-08-30
Document Information Extraction	Cloud Foundry	Free Service Plan	The Free service plan is now available for Document Information Extraction. See Service Plans [page 84], Tutorials [page 108] and Free Tier Option and Trial Account Technical Constraints [page 289].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-08-30	2022-08-30

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-08-30	2022-08-30
Document Information Extraction	Cloud Foundry	Extracted Header Fields	You can now extract the following header fields from `paymentAdvice` documents: • *senderAddress* • *taxId* See Extracted Header Fields [page 291].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-08-04	2022-08-04
Document Information Extraction	Cloud Foundry	New Business Card Supported Language: Hebrew	Document Information Extraction supports now `businessCard` documents in Hebrew. See Business Card: Languages [page 93].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-08-04	2022-08-04
Document Information Extraction	Cloud Foundry	Accessibility Features	Documentation on Accessibility Features in Document Information Extraction [page 309] is now available.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-08-04	2022-08-04

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-08-04	2022-08-04
Document Information Extraction	Cloud Foundry	Technical Constraints	The maximum number of clients you can create in one API call has increased from 10 to 5000. The maximum number of schemas per client and templates per tenant has increased from 100 to 1000. See Technical Constraints [page 288].	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-06-23	2022-06-23
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-06-23	2022-06-23
Document Information Extraction	Cloud Foundry	Handwriting Detection	The handwriting detection feature is now available. For now, it detects only handwriting in English. See Optical Character Recognition (OCR): Best Practices [page 272].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-06-23	2022-06-23

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Barcode Supported Countries/Regions and Extracted Fields for Invoice Documents	It's now available the list of supported countries/regions and extracted fields for barcodes in Invoice: Languages and Countries/Regions [page 94] documents.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-06-03	2022-06-03
Document Information Extraction	Cloud Foundry	New Supported Countries/Regions for Invoice Documents	Document Information Extraction supports now the following new countries/regions for Invoice: Languages and Countries/Regions [page 94] documents: • Austria • Belgium • Czech Republic • Denmark • Finland • Norway • Poland • Portugal • Slovakia • Slovenia • Sweden ⓘ Note To support the new languages, the machine learning models have been extended. Consequently, predictions (field extractions and corresponding confidence scores) may differ from previous releases.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-06-03	2022-06-03

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-06-03	2022-06-03
Document Information Extraction	Cloud Foundry	Document API	You can now see all matched enrichment data records in the Get Result [page 146] response.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-05-06	2022-05-06
Document Information Extraction	Cloud Foundry	Enrichment Data API	The Create Data Activation [page 187] endpoint now includes the optional parameters `type` and `subtype`.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-05-06	2022-05-06
Document Information Extraction	Cloud Foundry	Deskew	The service now automatically rotates document images to compensate for skewing. See Supported Document Types and File Formats [page 91].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-05-06	2022-05-06

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-05-06	2022-05-06
Document Information Extraction	Cloud Foundry	Document API	The Upload Document [page 135] endpoint now includes a schemaId parameter. This parameter is required in payloads that include templateId.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-04-22	2022-04-22
Document Information Extraction	Cloud Foundry	Enrichment Data API	You can now use variants to create multiple versions of the same data record. See Create Enrichment Data [page 175], Data Variants [page 181] and Data Duplicates [page 182].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-04-22	2022-04-22
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-04-22	2022-04-22

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Template	You can now use templates to extract multiple tables from the same page, provided the tables all have a standard structure and the same table headers. See General Recommendations and Limitations [page 279].	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-03-31	2022-03-31
Document Information Extraction	Cloud Foundry	Global Accounts	You can now move subaccounts between your global accounts. See Initial Setup [page 103].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-03-31	2022-03-31
Document Information Extraction	Cloud Foundry	Trial Account Technical Constraints	The Free Tier Option and Trial Account Technical Constraints [page 289] documentation has been updated.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-03-31	2022-03-31
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-03-31	2022-03-31

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Support for Multiple Service Instances	If you create more than one service instance, the Document Information Extraction UI now allows you to change between instances. See Subscribing to the Document Information Extraction UI [page 243].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-03-17	2022-03-17
Document Information Extraction	Cloud Foundry	Document Feature	You can now select folders containing multiple documents for upload. The Document Information Extraction UI now displays thumbnails of documents and allows you to rename documents before uploading them. See Add Document [page 251].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-03-17	2022-03-17
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements. Metering and pricing details for the Compute Hours for Base Edition [page 87] have been updated.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-03-17	2022-03-17
Document Information Extraction	Cloud Foundry	Document Extraction Results	You can now download extraction values before and after you edit and save them. See View and Edit Extraction Results [page 252].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-02-03	2022-02-03

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Document Extraction Results	You can now view the raw values for extraction results. Raw values are the original field values before postprocessing, which can differ from the corresponding extraction results. See View and Edit Extraction Results [page 252].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-02-03	2022-02-03
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code and stability improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-02-03	2022-02-03
Document Information Extraction	Cloud Foundry	SAP Schemas	The SAP schemas for standard document types now have the status ACTIVE. As a result, you no longer need to create copies of these schemas before using them to upload documents or create templates. See Schema Configuration [page 256].	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-01-18	2022-01-18
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-01-18	2022-01-18

Technical Component	Environment	Title	Description	Action	Lifecycle	Type	Line of Business	Modular Business Process	Product	Latest Revision	Available as of
Document Information Extraction	Cloud Foundry	Enrichment Data API	The new Enrichment Data API endpoint List Data-Persistence Jobs [page 182] is now available. The new enrichment data entity type Product [page 180] is now available.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-01-10	2022-01-10
Document Information Extraction	Cloud Foundry	Configuration API	The `performPIICheck` sub-configuration is now available. See Create Configuration [page 122].	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-01-10	2022-01-10
Document Information Extraction	Cloud Foundry	Mass Deletion of Documents	The Document [page 250] feature has been enhanced: you can now select multiple documents for simultaneous deletion.	Info only	General Availability	New	Technology	Not applicable	SAP Business Technology Platform	2022-01-10	2022-01-10
Document Information Extraction	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	General Availability	Changed	Technology	Not applicable	SAP Business Technology Platform	2022-01-10	2022-01-10

2.3   2021 What's New for Document Information Extraction (Archive)

Technical Component	Capability	Environment	Title	Description	Action	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	Changed	2021-12-06
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	Changed	2021-11-23
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Template Feature	Improved template extraction results for header fields in multipage documents. See Template [page 266].	Info only	Changed	2021-11-23
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Using the Document Information Extraction UI	Documentation updated: now includes requirement to use a schema when creating templates based on document extraction results. See Document [page 250] and Template [page 266].	Info only	Changed	2021-11-23

Technical Component	Capability	Environment	Title	Description	Action	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Tutorials	The following tutorial missions are now available for Document Information Extraction: • Shape Machine Learning to Process Standard Business Documents • Shape Machine Learning to Process Custom Business Documents See Tutorials [page 108].	Info only	New	2021-11-23
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	Changed	2021-11-05
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Enrichment Data API	The matching accuracy for the `bankAccount` `businessEntity` key has been improved. See BusinessEntity [page 178] and Data Enrichment: Best Practices [page 284].	Info only	Changed	2021-11-05
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Technical Constraints	The 3510 x 3510 pixels maximum limit for the file size of single-page JPEG, PNG and TIFF documents has been removed. You can now upload to the service documents with any resolution as long as the file size is not higher than 50 MB. See Optical Character Recognition (OCR): Best Practices [page 272] and Technical Constraints [page 288].	Info only	Deleted	2021-11-05

Technical Component	Capability	Environment	Title	Description	Action	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	Changed	2021-10-15
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	New AWS Region	Document Information Extraction is now available in the AWS region Australia (Sydney).	Info only	New	2021-10-15
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Support to Business Card Documents	Document Information Extraction supports now, at API level only, businessCard as one of the standard document types. See Supported Document Types and File Formats [page 91], Supported Languages and Countries/Regions [page 93], and Extracted Header Fields [page 291].	Info only	New	2021-10-15
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	Changed	2021-09-30

Tech-nical Com-po-nent	Capa-bility	Envi-ron-ment	Title	Description	Action	Type	Avail-able as of
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Role Collec-tions	The role collection `Document_Information_Extraction_UI_Admin_ User` is now available. This new collection provides access to all the features of the UI application. See Role Collections [page 245].	Info only	New	2021-09-30
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Best Practi-ces	Best practices covering all stages of processing documents in the Document Information Extraction UI are now available. See Document: Best Practices [page 283], Template: Best Practices [page 278], and Schema Configuration: Best Prac-tices [page 273].	Info only	New	2021-09-30
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Overall Im-prove-ments	There have been several code improvements.	Info only	Chang ed	2021-09-10
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Tem-plate Fea-ture	You can now: • Import and export templates • Create templates from extracted documents See Template [page 266].	Info only	New	2021-09-10

Technical Component	Capability	Environment	Title	Description	Action	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Supported File Formats	Single-page document files in TIFF format are now supported. See Supported Document Types and File Formats [page 91] and Technical Constraints [page 288].	Info only	New	2021-09-10
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	Changed	2021-08-31
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Technical Constraints	The technical constraints for the number of schemas are now available. See Technical Constraints [page 288] and Free Tier Option and Trial Account Technical Constraints [page 289].	Info only	Changed	2021-08-31
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	Changed	2021-08-12

Technical Component	Capability	Environment	Title	Description	Action	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Document API	The Get Result [page 146] endpoint returns now two new response fields: • `languageCodes` • `pageCount`	Info only	New	2021-08-12
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Configuration API	All Configuration API [page 122] keys have now tenant scope by default. Service instance scope is now also available for the `dataFeedbackCollection` and `documentRetentionTimeDays` keys. The `documentRetentionTimeDays` configuration key is now available. See Create Configuration [page 122].	Info only	New	2021-08-12
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Python Client Library	A Python client library is now available for Document Information Extraction. It provides easy access to the REST API, UI application, and facilitates the service onboarding process. Go to Python Client Library ↗ .	Info only	New	2021-07-26
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Barcode Header Field	Decoded information is now available for *barcode* fields from India invoices. See Extracted Header Fields [page 291] and Barcode Header Field in Invoice Documents [page 298].	Info only	New	2021-07-26

Technical Component	Capability	Environment	Title	Description	Action	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Template Feature	The Template [page 266] feature is now also available to all SAP BTP Trial users. See Free Tier Option and Trial Account Technical Constraints [page 289].	Info only	Changed	2021-07-26
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements. The Service Guide documentation has been updated: • Capabilities API [page 111] • Save Ground Truth [page 162] • Extracted Header Fields [page 291] • Extracted Line Items [page 300]	Info only	Changed	2021-07-26
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Template Feature	The Template [page 266] feature is now generally available to all Document Information Extraction UI application users.	Info only	New	2021-07-20
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Extension Capabilities Service Plan	The new Compute Hours for Base Edition [page 87] service plan is now available.	Info only	New	2021-07-20

Technical Component	Capability	Environment	Title	Description	Action	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	New	2021-07-07
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Document API	You can now use the Get Document File [page 167] endpoint to get the original document file you uploaded to the service.	Info only	New	2021-07-07
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Security Guide	Auditing and logging information is now available in the Security [page 302].	Info only	New	2021-07-07
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Schema Feature and Support for Custom Documents and Fields	The Schema Configuration [page 256] feature is now available in the Document Information Extraction UI application. Document Information Extraction supports now `custom` documents and fields. See Supported Document Types and File Formats [page 91].	Info only	New	2021-06-28

Technical Component	Capability	Environment	Title	Description	Action	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Support for Purchase Order Documents	Document Information Extraction supports now `purchaseOrder` documents for all users. The list of line items you can extract from `purchaseOrder` documents has been updated. See Extracted Line Items [page 300]. See also Supported Document Types and File Formats [page 91] and Extracted Header Fields [page 291].	Info only	Changed	2021-06-28
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Configuration API	The `dataFeedbackCollection` Configuration API [page 122] key is now available.	Info only	Changed	2021-06-28
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Template API (Beta)	The Template API (Beta) and its endpoints are no longer exposed to users at API level. The Template [page 266] feature remains available from the Document Information Extraction UI application.	Info only	Changed	2021-06-28
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Security Guide	The Security [page 302] documentation has been updated.	Info only	Changed	2021-06-28

Technical Component	Capability	Environment	Title	Description	Action	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	Changed	2021-05-24
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Barcode Header Field	The *barcode* header field can now be extracted from Ticket-BAI invoices for the three Basque provincial councils (Álava, Vizcaya and Guipúzcoa) and the Basque government. See Extracted Header Fields [page 291] and Barcode Header Field in Invoice Documents [page 298].	Info only	New	2021-05-24
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	Changed	2021-05-05
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Barcode Header Field	The *barcode* header field can now be extracted from: • Brazil PIX (instant payments) • Argentina, Colombia and Uruguay invoices See Extracted Header Fields [page 291] and Barcode Header Field in Invoice Documents [page 298].	Info only	New	2021-05-05

Tech-nical Com-po-nent	Capa-bility	Envi-ron-ment	Title	Description	Action	Type	Avail-able as of
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Overall Im-prove-ments	There have been several code improvements.	Info only	Chang ed	2021-03-29
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Sup-port for Fac-tur-X and ZUG-FeRD Stand-ards	Document Information Extraction supports now the Factur-X and ZUGFeRD standards (all versions) for e-invoice docu-ment files in PDF and XML hybrid format. See Supported Document Types and File Formats [page 91].	Info only	New	2021-03-29
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Docu ment Inform ation Extrac tion UI	The Document Information Extraction UI application now fea-tures: • Activation and deactivation of templates. See Template [page 266]. • Field level confidence visualization. See Document [page 250]. • Web Assistant	Info only	New	2021-03-29
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Tem-plate API (Beta)	The following Template API (Beta) endpoints are now availa-ble: • Activate Template (Beta) • Deactivate Template (Beta)	Info only	New	2021-03-29

Technical Component	Capability	Environment	Title	Description	Action	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Data Feedback Collection for Model Improvement	The data feedback collection feature is now available. See Get Result [page 146] and Confirm Document [page 165].	Info only	New	2021-03-29
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	Changed	2021-03-22
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Supported Languages and Countries/Regions	The list of supported countries/regions for `purchaseOrder` (controlled availability) documents, and the list of supported languages for the Template API (Beta) and the Document Information Extraction UI Template (Beta) feature are now available. See Supported Languages and Countries/Regions [page 93].	Info only	Changed	2021-03-22
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Barcode Header Field	Barcode header field extraction has been improved. See Extracted Header Fields [page 291] and Barcode Header Field in Invoice Documents [page 298].	Info only	Changed	2021-03-22

Technical Component	Capability	Environment	Title	Description	Action	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Info only	Changed	2021-03-01
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Enrichment Data API	You can now set data activation to manual, instead of using the default automatic refresh of enrichment data, that takes place every 4 hours. See Create Data Activation [page 187] and Get Data Activation Details [page 188].	Info only	Changed	2021-03-01
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Configuration API	The Configuration API [page 122] is now available.	Info only	New	2021-03-01
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Identifier API	The Identifier API (Deprecated) [page 117] is now available.	Info only	New	2021-03-01

Tech-nical Com-po-nent	Capa-bility	Envi-ron-ment	Title	Description	Action	Type	Avail-able as of
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Sup-ported Docu-ment Types and File For-mats	Document Information Extraction supports now `paymentAdvice` document files in Excel format. See Sup-ported Document Types and File Formats [page 91].	Info only	New	2021-03-01
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Docu-ment API	The `rawValue` response field is now available for the Get Result [page 146] endpoint.	Info only	Chang ed	2021-02-15
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	SAP API Busi-ness Hub	Document Information Extraction is now available in the SAP API Business Hub. See Document Information Extraction .	Info only	New	2021-02-15
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	En-rich-ment Data API	You can now delete large numbers of data records for all clients per data type (`employee` or `businessEntity`). See Delete Enrichment Data (Asynchronous) [page 191].	Info only	Chang ed	2021-02-01

Tech-nical Com-po-nent	Capa-bility	Envi-ron-ment	Title	Description	Action	Type	Avail-able as of
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Overall Im-prove-ments	There have been several code improvements.	Info only	Chang ed	2021-02-01
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Tem-plate (Beta) Fea-ture	The Document Information Extraction UI Template [page 266] feature has been updated. See Add Documents and Ac-tivate/Deactivate Template [page 267]. The role collection `Document_Information_Extraction_UI_Templa tes_Admin` is now available. See Subscribing to the Docu-ment Information Extraction UI [page 243].	Info only	Chang ed	2021-01-18
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Overall Im-prove-ments	There have been several code improvements.	Info only	Chang ed	2021-01-18
Docum ent Inform ation Extract ion	Exten-sion Suite - Devel-op-ment Effi-ciency	Cloud Foun-dry	Ex-tracte d Heade r Fields	The list of header fields you can extract from `purchaseOrder` documents has been updated. See Extracted Header Fields [page 291].	Info only	Chang ed	2021-01-04

2.4 2020 What's New for Document Information Extraction (Archive)

Technical Component	Capability	Environment	Title	Description	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Changed	2020-12-21
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Template (Beta) Feature	The Document Information Extraction UI Template [page 266] feature supports now `purchaseOrder` documents (only for a previously selected group of beta customers). See Using the Document Information Extraction UI [page 243].	New	2020-12-21
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Changed	2020-12-03
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	New SAP Cloud Platform Cockpit Booster	You can now use the *Set up account for Document Information Extraction* booster to automate the onboarding steps on the SAP Cloud Platform cockpit, and quickly consume the service and its UI application. See Initial Setup [page 103] and Subscribing to the Document Information Extraction UI [page 243].	New	2020-11-20
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	New Beta Features	• Document Information Extraction supports now `purchaseOrder` documents (only for a previously selected group of beta customers). See Supported Document Types and File Formats [page 91], Extracted Header Fields [page 291] and Extracted Line Items [page 300]. • The Template [page 266] feature is now available (only for a previously selected group of beta customers) in the Document Information Extraction UI. See Using the Document Information Extraction UI [page 243].	New	2020-11-20

Technical Component	Capability	Environment	Title	Description	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	• There have been several code improvements. • The Feature Scope Description for Document Information Extraction has been updated. • The Technical Constraints [page 288] have been updated. • The Document Information Extraction Tutorials [page 108] have been updated.	Changed	2020-11-20
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	New AWS Region	Document Information Extraction is now available in the AWS region Europe (Frankfurt) EU-ONLY (access from Europe only).	New	2020-10-27
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	• There have been several code improvements. • The *discount* and *dueDate* header fields can now be extracted from invoices. See Extracted Header Fields [page 291]. • To get better extraction and enrichment results with Document Information Extraction, see Optical Character Recognition (OCR): Best Practices [page 272].	Changed	2020-10-27
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Metering and Pricing	A new service plan is available for Document Information Extraction. See Metering and Pricing [page 86].	New	2020-10-21
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements: • The *barcode* header field can now be extracted from India invoices. See Extracted Header Fields [page 291]. • The new `returnNullValues` request parameter is now available for the **Get Result** endpoint. See Get Result [page 146].	Changed	2020-10-16

Technical Component	Capability	Environment	Title	Description	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Extracted Line Items	The *unitOfMeasure* line item can now be extracted from invoices. See Extracted Line Items [page 300].	Changed	2020-10-05
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	UI Application	The Document Information Extraction UI is now generally available to all SAP Cloud Platform customers. See Using the Document Information Extraction UI [page 243].	New	2020-10-05
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Supported Document Types and File Formats	The Service Guide documentation has been updated with a new section: Supported Document Types and File Formats [page 91].	New	2020-09-16
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements: • The *barcode* header field can now be extracted from invoices. See Extracted Header Fields [page 291]. • The `fileType` response field is now available for the Get Result [page 146] endpoint.	Changed	2020-09-16
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Using the Document Information Extraction UI (Beta)	A new version of the Document Information Extraction UI is now available (only for a previously selected group of beta customers). See details on the possible document statuses and the *Confirm* document functionality in Using the Key Features of the Document Information Extraction UI [page 246].	Changed	2020-08-28

Technical Component	Capability	Environment	Title	Description	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Changed	2020-08-28
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Document API	The `clientId` request parameter is no longer needed to send a request to the following Document API [page 134] endpoints: • Get Result [page 146] • Save Ground Truth [page 162] • Get All Pages Text [page 168] • Get Single Page Text [page 170] • Get Request Options [page 172]	Changed	2020-08-17
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Changed	2020-08-17
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	New SAP Cloud Platform Trial Cockpit Booster	You can now use the *Set up account for Document Information Extraction* booster to automatically create your Document Information Extraction service key on SAP Cloud Platform Trial. Follow the steps described in the tutorial Set Up Account for Document Information Extraction .	New	2020-08-17
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	New AWS Region	Document Information Extraction is now available in the AWS region US East (VA).	New	2020-07-31

Technical Component	Capability	Environment	Title	Description	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	New document file formats for `paymentAdvice`	Single-page PNG and JPEG `paymentAdvice` files are now supported. See Upload Document [page 135] and Technical Constraints [page 288].	New	2020-07-31
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Using the Document Information Extraction UI (Beta)	A new version of the Document Information Extraction UI is now available (only for a previously selected group of beta customers). See Using the Document Information Extraction UI [page 243].	Changed	2020-07-31
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements.	Changed	2020-07-31
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code and usability improvements: • Enrichment data upload performance. See Create Enrichment Data [page 175]. • Document confirmation feature. See the new Document API endpoint Confirm Document [page 165].	Changed	2020-07-14
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code improvements: • The *deliveryDate*, *paymentTerms* and *senderBankAccount* header fields can now be extracted from invoices. See Extracted Header Fields [page 291]. • The list of supported character types for the IDs of clients, enrichment data records, system and company codes has been updated. See Technical Constraints [page 288].	Changed	2020-06-15

Technical Component	Capability	Environment	Title	Description	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code and usability improvements: • Single-page PNG and JPEG invoice files are now supported. See Upload Document [page 135] and Technical Constraints [page 288]. • New Document API [page 134] endpoints are now available. • The Enrichment Data API [page 174] endpoints have also been updated. Delete Enrichment Data (Asynchronous) [page 191] is now available. • The *deliveryNoteNumber* header field can now be extracted from invoices. See Extracted Header Fields [page 291]. • You can now use the Capabilities API [page 111] to get the list of document fields and enrichment data you can process by document type.	Changed	2020-06-02
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	New Beta Features	The following beta features are now available (only for a previously selected group of beta customers): • Template-based information extraction. See Template API (Beta) and Technical Constraints [page 288]. • Document Information Extraction UI. See Using the Document Information Extraction UI [page 243].	New	2020-06-02
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several code and usability improvements: • Higher model accuracy • The Supported Languages and Countries/Regions [page 93] list has been updated • The tutorial mission Use Machine Learning to Enrich Data Extracted from Documents is now available. See Tutorials [page 108].	Changed	2020-05-18
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	New Notifications Functionality	The Notifications [page 236] functionality is now available.	New	2020-05-18

Technical Component	Capability	Environment	Title	Description	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several stability and usability improvements, including the model accuracy. The Service Guide documentation has been updated: • Technical Constraints [page 288] • Free Tier Option and Trial Account Technical Constraints [page 289]	Changed	2020-04-20
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Overall Improvements	There have been several stability and usability improvements: • Some field value types have been updated. See Capabilities API [page 111] • The enrichment parameter top property has now a maximum possible value of 50. See Enrichment Parameter [page 141]. • Now, if no value is detected for fields in header or line items, they do not appear in the response JSON file. See Get Result [page 146].	Changed	2020-03-30
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	API Reference	The API Reference [page 109] documentation has been updated with the following new sections: • Get Access Token [page 110] • Capabilities API [page 111] • Technical Constraints [page 288]	Changed	2020-03-30
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Tutorials	A new tutorial mission is now available for Document Information Extraction. See Use Machine Learning to Process Business Documents.	New	2020-03-02
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Client API	The new clientIdStartsWith request parameter is now available for the **Get Client** endpoint. See Get Client [page 115] .	New	2020-03-02

2.5 2019 What's New for Document Information Extraction (Archive)

Technical Component	Capability	Environment	Title	Description	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	New AWS Region	Document Information Extraction is now available in the AWS region Japan (Tokyo).	New	2019-12-19
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Trial Account	You can now try out Document Information Extraction on SAP Cloud Platform Trial. See Get a Trial Account.	New	2019-12-05
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	API Reference	• Enrichment Data API documentation is now available. See Enrichment Data API [page 174]. • Document API documentation has also been updated. See Document API [page 134]. • The *documentNumber*, *documentDate*, *discountAmount*, *deductionAmount*, and *grossAmount* fields can now be extracted from line items. See Extracted Line Items [page 300].	Changed	2019-11-04
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Getting Support	CA-ML-BDP is now the BCP component for Document Information Extraction. See Getting Support [page 310].	Changed	2019-11-04

Technical Component	Capability	Environment	Title	Description	Type	Available as of
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Security Guide	The Security Guide has been updated with Enrichment Data API details. See Security [page 302].	Changed	2019-11-04
Document Information Extraction	Extension Suite - Development Efficiency	Cloud Foundry	Troubleshooting	The Troubleshooting section is now available. See Troubleshooting [page 311].	New	2019-11-04

3   Concepts

See a glossary of definitions for artificial intelligence (AI) and machine learning (ML), and Document Information Extraction concepts in AI & ML Glossary. In the third column *Filter*, select **Document Information Extraction**.

4 Service Plans

Learn more about the different types of service plans for Document Information Extraction.

Document Information Extraction provides different types of service plans. The type you choose determines pricing, conditions of use, resources, available services, and hosts.

It depends on your use case whether you choose a free or a paid service plan. If you plan to use your global account in productive mode, you must purchase a paid enterprise account. It's important that you're aware of the differences when you're planning and setting up your account model. See Initial Setup [page 103].

The following service plans are currently available:

- For enterprise and trial accounts: **Base Edition** (`blocks_of_100`)
- For enterprise accounts: **Premium Edition** (`premium_edition`)
- For enterprise accounts: **Free** (`free`)

For more details about the available service plans, see the following table:

Service Plan	Details	Account Type
Base Edition (`blocks_of_100`)	- Base Edition service plan that includes all core features but doesn't include document information extraction using generative AI. - Service plan intended for productive usage. - Inference requests in blocks of 100 documents and compute hours. - You can upload to the service up to 2000 documents per hour per tenant (each document can have up to 100 pages). See Metering and Pricing [page 86] and Technical Constraints [page 288].	Enterprise
Base Edition (`blocks_of_100`)	- Service plan intended for personal exploration. Access is open to everyone after registration. - It includes document information extraction using **generative AI**. - You can upload to the service up to 50 document pages per tenant in a rolling period of 30 days. See Free Tier Option and Trial Account Technical Constraints [page 289].	Trial

Service Plan	Details	Account Type
Free (`free`)	• Service plan intended for development and try-out purposes on your enterprise account. • It doesn't include document information extraction using generative AI. • You can upload to the service up to 50 document pages per tenant in a rolling period of 30 days. See Free Tier Option and Trial Account Technical Constraints [page 289] and the tutorial Get an Account on SAP BTP to Try Out Free Tier Service Plans .	Enterprise
Premium Edition (`premium_edition`)	• Premium edition service plan that includes document information extraction using **generative AI**. • Service plan intended for productive usage. • Inference requests in blocks of 100 documents. • You can upload to the service up to 2000 documents per hour per tenant (each document can have up to 100 pages). See Metering and Pricing [page 86] and Technical Constraints [page 288].	Enterprise

> → Remember
>
> • If you first activated the **Free** service plan, you can update the same service instance to switch to **Base Edition** or **Premium Edition** for enterprise accounts.
> • Both metadata and transaction data are transferred to **Base Edition** or **Premium Edition** for enterprise accounts when you switch from **Free** to **Base Edition** or **Premium Edition**.
> • If you don't want **Free** and **Base Edition** or **Premium Edition** data to be combined together, you can split them by subscribing to the service plans in separate subaccounts.

5 Metering and Pricing

Learn more about the different types of metering and pricing for Document Information Extraction by service plan.

> → Tip
>
> The metering and pricing details listed here are relevant only to users of the service plans **Base Edition** (`blocks_of_100`) and **Premium Edition** (`premium_edition`) for enterprise accounts. See Service Plans [page 84].

The service plan **Document Information Extraction, base edition** (`blocks_of_100`) is metered based on the following metrics:

- Blocks of 100 Documents for Base Edition [page 86]
- Compute Hours for Base Edition [page 87]

The service plan **Document Information Extraction, premium edition** (`premium_edition`) is metered based on the following metrics:

- Blocks of 100 Documents for Premium Edition [page 89]

> → Tip
>
> Use the pricing estimator tool.

Related Information

SAP Discovery Center
SAP Business Technology Platform Service Description Guide

5.1 Blocks of 100 Documents for Base Edition

Usage Metric

The service plan **Document Information Extraction, base edition** (`blocks_of_100`) is metered based on the usage of documents defined as unique records processed by the cloud service. One document can consist of maximum 3 pages. If a document consists of more than 3 pages, each additional 3 pages are charged as an additional document.

Block Size

1 block = 100 documents. The final price is a sum of the number of documents uploaded to the service.

Basic Service

> ⚠ Caution
>
> The price rates listed below might be outdated. Find updated price rates in the *Pricing* tab of the SAP Discovery Center ⛅.

Document Information Extraction does not allow users to train and deploy customizable models. For this service, the number of inference requests is relevant for the charged amount.

Metric	Tiers	Block Price per Month
Blocks of 100 documents	1 to 300 blocks	EUR 20.00
1 document = 3 pages	301 to 600 blocks	EUR 17.00
	More than 601 blocks	EUR 14.00

Example

Cost for 7 blocks = 7 * EUR 20.00 = EUR 140.00.

Cost for 310 blocks = 310 * EUR 17.00 = EUR 5,270.

Cost for 610 blocks = 610 * EUR 14.00 = EUR 8,540.

5.2 Compute Hours for Base Edition

Usage Metric

The service plan **Document Information Extraction, base edition** (`blocks_of_100`) is also metered based on consumed compute hours defined as one hour, or portion thereof, consumed by the cloud service to process one or more documents with a custom model.

Piece Size

> **⚠ Caution**
>
> The price rate listed below might be outdated. Find updated price rates in the *Pricing* tab of the SAP Discovery Center🔗.

1 piece = compute hour. 1 compute hour = EUR 1.00.

1 template activation = 5 compute hours.

The costs are associated with the usage of templates. See Template API [page 219] and the Template [page 266] UI feature.

Example

> **ⓘ Note**
>
> The following calculation examples are based on current experiments. During the usage of the service, the exact usage numbers can vary slightly.

Basic Service Calculation

- Metric = compute hours (usage of templates)
- 1 compute hour = EUR 1.00
- 1 template activation = 5 compute hours (EUR 5.00)
- 1 template transaction = 1 second

Number of Documents (per Month)	Estimated Processing Time in Compute Hours (per Month)	Estimated Costs in Euro (per Month)
500 documents	500 seconds = 1 compute hour	EUR 1.00
1000 documents	1000 seconds = 1 compute hour	EUR 1.00
5000 documents	5000 seconds = 2 compute hours	EUR 2.00
10,000 documents	10,000 seconds = 3 compute hours	EUR 3.00

Human Resources (HR) Onboarding Business Scenario Calculation

- HR Shared Service Center onboards 200 new employees each month.
- For each new employee, the HR agent needs to extract information from five different document types.
- The HR agent needs to:
  - Process 1000 documents per month
  - Create and activate five custom templates
  - Use each template 200 times per month

Cost Components	Consumption	Price
Document Upload	1000 documents (10 blocks of 100 documents)	EUR 200 per month
Template activation	5 templates * 5 compute hours	EUR 25.00 one time
Template usage	1000 (5 templates * 200 transactions) * 1 second = 0.3 compute hour (rounded up to 1 compute hour)	EUR 1.00 per month

In this example, the total cost is EUR 201.00 per month, and EUR 25.00 only once when the five templates are activated.

5.3 Blocks of 100 Documents for Premium Edition

Usage Metric

The service plan **Document Information Extraction, premium edition** (`premium_edition`) is metered based on the usage of documents defined as unique records processed by the cloud service.

One document can consist of maximum 1 page. If a document consists of more than 1 page, each additional page is charged as an additional document.

You can extract a maximum of 50 fields per document. If you extract more than 50 fields per document, every additional 50 fields are charged as an additional document. As a technical limit, you can add up to 500 header fields and line items per schema. For more information, see Technical Constraints [page 288].

Block Size

1 block = 100 documents. The final price is a sum of the number of documents uploaded to the service.

Basic Service

> ⚠ Caution
>
> The price rates listed below might be outdated. Find updated price rates in the *Pricing* tab of the SAP Discovery Center .

Document Information Extraction does not allow users to train and deploy customizable models. For this service, the number of inference requests is relevant for the charged amount.

Metric	Tiers	Price per Month
Blocks of 100 documents	1 to 5 blocks	EUR 300.00 (fixed price)
1 document = 1 page	More than 5 blocks	EUR 60.00 (block price)

Example

Cost for 1 block = EUR 300.00.

Cost for 3 blocks = EUR 300.00.

Cost for 5 blocks = EUR 300.00.

Cost for 10 blocks = 10 * EUR 60.00 = EUR 600.00.

6 Supported Document Types and File Formats

Document Types

Document Information Extraction supports the following document types as input:

- Standard document types: refer to document types for which SAP provides pre-trained machine learning models that allow out-of-the-box (without prior training) extraction of information based on default extractors, which are managed directly by SAP.

  - `businessCard`

    > ⓘ Note
    >
    > - For now, `businessCard` documents are **only** supported at API level.
    > - Be aware that `businessCard` documents with more than one contact person are **not** supported.
    > - For `businessCard` documents, the service extracts only the information (contact details) from the first page of any submitted document, but all pages are counted for metering purposes. Submit only single-page documents to avoid additional charges. See Metering and Pricing [page 86].

  - `invoice`
  - `paymentAdvice`
  - `purchaseOrder`

- Custom document types: refer to document types for which there are no pre-trained machine learning models that are managed by SAP. Use the Template [page 266] and Schema Configuration [page 256] features to extract information from `custom` documents that are different from the `standard` document types listed above. See also Schema API [page 193] and Template API [page 219].

File Formats

Document Information Extraction supports the following document file formats as input:

- Document files in PDF format
- Single-page document files in JPEG, PNG, and TIFF format
- Image files that include scene text in JPEG, PNG, and TIFF format
- E-invoice document files in PDF and XML hybrid format, and in Factur-X and ZUGFeRD standards (all versions)
- `paymentAdvice` document files in Excel format

ⓘ Note

- The endpoint Upload Document [page 135] accepts only multipart-encoded files with a file name and a content type.
- The file name should contain a file extension. For example: "invoice" only, without a file extension, is **not** a valid file name.
- The file name cannot be empty even if a file extension is provided. For example: ".pdf" is **not** a valid file name.

→ Tip

The Document Information Extraction service handles distorted and asymmetrical images with a rotation of multiples of 90 degrees. In addition, small rotations of up to 15 degrees are also handled by the service. In both cases, the images are deskewed automatically.

7 Supported Languages and Countries/ Regions

Explore the Document Information Extraction supported languages and countries/regions by document type and extraction approach.

The supported languages and countries/regions have been validated with Document Information Extraction. It is also possible to get similar accuracy results with documents in other languages and from other countries/ regions that use *Latin-1 (ISO-8859-1)* character sets.

If you want to try out Document Information Extraction to check if it fulfills your business needs, you can use a trial account to upload to the service a document in any language and from any country/region, and get the results following the tutorial mission Use Machine Learning to Process Business Documents .

7.1 Business Card: Languages

See the list of supported languages for `businessCard` documents.

> **⚠ Restriction**
>
> For now, `businessCard` documents are **only** supported at API level.

Language

Document Information Extraction supports the following languages for `businessCard` documents:

Language	Language Code
Chinese Simplified	`zh-Hans`
Chinese Traditional	`zh-Hant`
Dutch	`nl`

Language	Language Code
English	en
French	fr
German	de
Hebrew	he
Italian	it
Japanese	ja
Korean	ko
Polish	pl
Portuguese	pt
Russian	ru
Spanish	es

7.2 Invoice: Languages and Countries/Regions

See the list of supported languages and countries/regions for `invoice` documents. See also the supported countries/regions, and extracted fields for barcodes in `invoice` documents.

Language

Document Information Extraction supports the following languages for `invoice` documents:

Language	Language Code
Czech	cs
Danish	da
Dutch	nl
English	en
Finnish	fi
French	fr
German	de
Hungarian	hu
Italian	it
Japanese	jp

Language	Language Code
Norwegian	`no`
Polish	`pl`
Portuguese	`pt`
Romanian	`ro`
Slovak	`sk`
Slovenian	`sl`
Spanish	`es`
Swedish	`sv`
Turkish	`tr`

Country/Region

Document Information Extraction supports the following countries/regions for `invoice` documents:

- Australia
- Austria
- Belgium
- Canada
- Czech Republic
- Denmark
- Finland
- France
- Germany
- Hungary
- Italy
- Japan
- Mexico
- Netherlands
- New Zealand
- Norway
- Poland
- Portugal
- Romania
- Slovakia
- Slovenia
- Spain
- Sweden
- Switzerland

- Türkiye
- United Kingdom
- United States

> ⓘ Note
>
> Document Information Extraction does not guarantee to support all specific fields for the countries/ regions listed above, even if they are legally required in a country/region.

Barcode Country/Region and Extracted Fields

Document Information Extraction supports the following countries/regions, and extracted fields for barcodes in `invoice` documents:

Barcode Country/Region	Extracted Fields
Argentina	• *currencyCode* • *documentDate* • *documentNumber* • *grossAmount*
Basque	• *documentNumber* • *grossAmount*
Brazil	• *currencyCode* • *grossAmount* • *senderName*
China	• *documentDate* • *documentNumber* • *netAmount*
Colombia	• *documentDate* • *documentNumber* • *grossAmount* • *netAmount* • *receiverTaxId* • *taxAmount*
EPC QR code (European Payments Council Quick Response Code including Austria, Belgium, Finland, Germany, and Netherlands)	• *currencyCode* • *grossAmount* • *senderName*

Barcode Country/Region	Extracted Fields
India	• *documentDate* • *documentNumber* • *grossAmount* • *receiverTaxId* • *taxId*
Mexico	• *grossAmount* • *taxId*
Switzerland	• *currencyCode* • *documentNumber* • *grossAmount* • *senderAddress* • *senderBankAccount* • *senderName* • *receiverAddress* • *receiverName*
Uruguay	• *documentNumber* • *grossAmount*

> ⓘ **Note**
>
> The barcode supported countries/regions have been validated with Document Information Extraction. It is also possible to get similar accuracy results with documents from other countries/regions that contain the most common types of 1D and 2D barcodes as described in Barcode Header Field in Invoice Documents [page 298].

7.3 Payment Advice: Languages and Countries/Regions

See the list of supported languages and countries/regions for `paymentAdvice` documents.

Language

Document Information Extraction supports the following languages for `paymentAdvice` documents:

Language	Language Code
English	en

Language	Language Code
German	de

Country/Region

Document Information Extraction supports the following countries/regions for `paymentAdvice` documents:

- Germany
- United Kingdom

7.4   Purchase Order: Languages and Countries/Regions

See the list of supported languages and countries/regions for `purchaseOrder` documents.

Language

Document Information Extraction supports the following languages for `purchaseOrder` documents:

Language	Language Code
English	en
German	de

Country/Region

Document Information Extraction supports the following countries/regions for `purchaseOrder` documents:

- Germany
- United Kingdom
- United States

7.5 Extraction Using Template: Languages

See the list of languages supported when using a template to extract information from `custom` and `standard` documents.

> ⓘ Note
>
> When using templates to extract information from `standard` documents, the accuracy results are usually higher when you take into account the supported languages and countries/regions listed for Business Card: Languages [page 93], Invoice: Languages and Countries/Regions [page 94], Payment Advice: Languages and Countries/Regions [page 97], and Purchase Order: Languages and Countries/Regions [page 98] documents.

Language

Extraction using a template supports the following **Latin** languages:

Language	Language Code
Albanian	sq
Bosnian	bs
Catalan	ca
Croatian	hr
Czech	cs
Danish	da
Dutch	nl
English	en
Estonian	et
Finnish	fi
French	fr
Galician	gl
German	de
Hungarian	hu
Icelandic	is
Indonesian	id
Italian	it
Irish	ga

Language	Language Code
Latvian	`lv`
Lithuanian	`lt`
Malaysian	`ms`
Montenegrin	`cnr`
Norwegian	`no`
Polish	`pl`
Portuguese	`pt`
Serbian	`sr`
Slovak	`sk`
Slovenian	`sl`
Spanish	`es`
Swedish	`sv`
Turkish	`tr`
Welsh	`cy`

Extraction using a template supports the following **non-Latin** languages:

Language	Language Code
Arabic	`ar`
Chinese Simplified	`zh-Hans`
Chinese Traditional	`zh-Hant`
Greek	`el`
Hebrew	`he`
Japanese	`ja`
Korean	`ko`
Russian	`ru`
Thai	`th`

Related Information

7.6 Extraction Using Generative AI: Languages

See the list of languages supported when using generative AI to extract information from `custom` and `standard` documents.

> ⚠ **Restriction**
>
> Extraction using generative AI is available with the service plan **Document Information Extraction, premium edition** (`premium_edition`) only. See Service Plans [page 84] and Metering and Pricing [page 86].
>
> You can also use an SAP BTP trial account to try out extraction using generative AI. Follow the tutorial: Use Trial to Extract Information from Custom Documents with Generative AI and Document Information Extraction🔗.

Language

Extraction using generative AI supports the following **Latin** languages:

Language	Language Code
Albanian	sq
Bosnian	bs
Catalan	ca
Croatian	hr
Czech	cs
Danish	da
Dutch	nl
English	en
Estonian	et
Finnish	fi
French	fr
Galician	gl
German	de
Hungarian	hu
Icelandic	is
Indonesian	id
Italian	it

Language	Language Code
Irish	`ga`
Latvian	`lv`
Lithuanian	`lt`
Malaysian	`ms`
Montenegrin	`cnr`
Norwegian	`no`
Polish	`pl`
Portuguese	`pt`
Serbian	`sr`
Slovak	`sk`
Slovenian	`sl`
Spanish	`es`
Swedish	`sv`
Turkish	`tr`
Welsh	`cy`

Extraction using generative AI supports the following **non-Latin** languages:

Language	Language Code
Arabic	`ar`
Chinese Simplified	`zh-Hans`
Chinese Traditional	`zh-Hant`
Greek	`el`
Hebrew	`he`
Japanese	`ja`
Korean	`ko`
Russian	`ru`
Thai	`th`

Related Information

Add Fields to Schema Version [page 208]
Setup Types [page 261]
Extraction Using Generative AI: Best Practices [page 286]

8   Initial Setup

Get started with Document Information Extraction using the standard procedures for SAP BTP Cloud Foundry environment or Kyma environment.

> → Tip
>
> See Tutorials [page 108] to find out how to use a trial account or the free tier option for Document Information Extraction to try out the service.

Prerequisites

You have set up your global account and at least one subaccount on SAP BTP. For an overview of the required steps, see Getting Started in the Cloud Foundry Environment or Getting Started in the Kyma Environment.

> ⓘ Note
>
> Document Information Extraction allows you to move subaccounts between your global accounts. For more information, see Relationship Between Global Accounts, Subaccounts, and Directories [Feature Set B].

Related Information

Enabling the Service in the Cloud Foundry Environment [page 104]
Enabling the Service in the Kyma Environment [page 104]
Subscribing to the Document Information Extraction UI [page 243]

8.1 Enabling the Service in the Cloud Foundry Environment

Enable Document Information Extraction using the standard procedures for SAP BTP Cloud Foundry environment.

Context

> → Tip
>
> You can also use the booster *Set up account for Document Information Extraction* to automate the steps described below on the SAP BTP cockpit. See Boosters and the tutorials:
>
> - Use Free Tier to Set Up Account for Document Information Extraction and Get Service Key
> - Use Free Tier to Set Up Account for Document Information Extraction and Go to Application

Procedure

1. Create a service instance in the Cloud Foundry environment. See Creating Service Instances.
2. You can then bind the service instance to your application, or you can create a service key to communicate directly with the service instance. See Binding Service Instances to Applications and Creating Service Keys.

8.2 Enabling the Service in the Kyma Environment

Enable Document Information Extraction using the standard procedures for Kyma environment.

Procedure

1. Create a service instance in the Kyma environment.
2. You can then bind the service instance to your application, or you can create a service key to communicate directly with the service instance. See Using SAP BTP Services in the Kyma Environment.

9 Enable X.509 Authentication

Find out how to enable your service instance for authentication with an X.509 client certificate.

The Document Information Extraction service supports X.509 authentication with the certificates managed either by the SAP Authorization and Trust Management service or self-managed. The authentication with an X.509 client certificate is enabled for every service instance by default.

Create Service Key or Service Binding Additional Parameters

To use X.509 secrets, you need to set additional parameters when you create your service key or service binding. We support the following two scenarios:

- The SAP Authorization and Trust Management service generates certificates for you. In this case, the parameters you need to set when you create your service key or service binding are the following in `JSON` format:

```
{
    "xsuaa":{
        "credential-type":"x509",
        "x509":{
            "key-length":2048,
            "validity":8,
            "validity-type":"DAYS"
        }
    }
}
```

For a detailed description of the parameters, see Parameters for X.509 Certificates Managed by SAP Authorization and Trust Management Service.

- You already have your own public key infrastructure (PKI), with certificates issued from one of the trusted Certificate Authorities (CAs). In this case, the parameters you need to set when you create your service key or service binding are the following in `JSON` format::

```
{
    "xsuaa":{
        "credential-type":"x509",
        "x509":{
            "certificate":"-----BEGIN CERTIFICATE-----...-----END
CERTIFICATE-----",
            "ensure-uniqueness":false,
            "certificate-pinning":true,
            "hide-certificate":true
        }
    }
}
```

For a detailed description of the parameters, see Parameters for Self-Managed X.509 Certificates. See also Trusted Certificate Authentication.

Get an Authorization Token with X.509 Certificate

To get an authorization token using an X.509 certificate, use "certurl". In the scenario of already generated certificates, also use "key" and "certificate" from the service key.

Example of a request using `curl`**:**

```
curl --cert <path to certificate.pem> --key <path to
key.pem> --request POST <value of "uaa.certurl">/oauth/token -d
'grant_type=client_credentials&client_id=<Value of "uaa.clientid">'
```

See also the blog post: X.509 certificate-based authentication(mTLS) – Generating X.509 certificates of BTP managed services.

Restrict your Service Instance to X.509 Certificate Authentication Only (Optional)

To restrict your instance to support **only** X.509 certificate authentication and don't allow any service keys that use different authentication methods, in the **New Instance or Subscription** wizard, enter in **Parameters** the following instance parameters in `JSON` format:

```
{
    "xs-security":{
        "xsappname":"<app-name>",
        "oauth2-configuration":{
            "credential-types":[
                "x509"
            ]
        }
    }
}
```

> ⓘ Note
>
> In the sample code, "<app-name>" is a name of your choice.

10 Run the Service in a Multitenant Application

Find out how to run Document Information Extraction in a multitenant application.

In the Cloud Foundry environment, you can develop and run multitenant applications, and share them with multiple consumers simultaneously on SAP BTP.

Document Information Extraction supports this scenario and can be declared as a dependency of a multitenant application. This means that Document Information Extraction gets provisioned automatically for every consumer that subscribes to the multitenant application. Different consumers are independently provisioned and data from these consumers is isolated inside Document Information Extraction.

> → Tip
>
> See Developing Multitenant Applications in the Cloud Foundry Environment for more details on how to declare Document Information Extraction as a dependency of a multitenant application using the SAP SaaS Provisioning service.

11 Tutorials

Follow our tutorials to get familiar with the Document Information Extraction UI application, APIs, and functionalities.

Tutorial Missions	Description
Use Generative AI to Process Business Documents📎	Find out how to use the SAP Business Technology Platform service Document Information Extraction with generative AI to automate the extraction of information from any type of document using large language models (LLMs).
Use Machine Learning to Process Business Documents📎	Try out the Document Information Extraction Trial UI to process business documents that have content in headers and tables.
Use Machine Learning to Extract Information from Business Documents and Enrich Data📎	Process business documents that have content in headers and tables, and enrich the information extracted with your own master data records, using machine learning and `Swagger UI`.
Shape Machine Learning to Process Standard Business Documents📎	Create your own header and line item fields, and edit extraction results for documents associated with templates to automate the extraction of information from standard business documents such as invoices and purchase orders.
Shape Machine Learning to Process Custom Documents📎	Create your own header and line item fields, and edit extraction results for documents associated with templates to automate the extraction of information from custom documents (not supported out of the box) such as résumés and power of attorney.

> → Tip
>
> See also the following onboarding tutorials that use the free tier option for Document Information Extraction:
>
> - Use Free Tier to Set Up Account for Document Information Extraction and Get Service Key📎
> - Use Free Tier to Set Up Account for Document Information Extraction and Go to Application📎

Related Information

Tutorial Navigator📎

12 Development

Explore the sections listed below to get started with the Document Information Extraction APIs and the Notifications feature.

- API Reference [page 109]
- Notifications [page 236]

12.1 API Reference

Explore the Document Information Extraction APIs.

Before using the Document Information Extraction APIs listed below, you need to retrieve your OAuth access token as described in Get Access Token [page 110].

- Capabilities API [page 111]
- Client API [page 114]
- Identifier API (Deprecated) [page 117]
- Configuration API [page 122]
- Document API [page 134]
- Enrichment Data API [page 174]
- Schema API [page 193]
- Template API [page 219]

To display the comprehensive specification of the Document Information Extraction APIs in `Swagger UI`, add the URL path extension `/document-information-extraction/v1` to the Document Information Extraction base URL (that is, the url value from outside the uaa section of your service key).

Related Information

12.1.1 Get Access Token

Retrieve your OAuth access token, which will grant you access to the Document Information Extraction APIs.

> ⓘ Note
>
> The token is valid for 12 hours. After that, you need to generate a new one.

> → Tip
>
> Alternatively, you can follow the steps in this tutorial to Get OAuth Access Token for Document Information Extraction via Web Browser👉.

Request

Base URL: url value from inside the uaa section of the service key

URL Path: `/oauth/token`

HTTP Method: *POST*

Request Headers

Header	Required	Values
`Content-Type`	Yes	`<application/x-www-form-urlencoded>`

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`client_id`	Yes	String	*authorization header*	The `clientid` value from the service key.
`client_secret`	Yes	String	*authorization header*	The `clientsecret` value from the service key.
`grant_type`	Yes	String	*query*	Token grant type. Set it to `client_credentials`.
`response_type`	Yes	String	*query*	Token response type. Set it to `token`.

Response

The response is given as a status (200 or 401). See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "access_token": "<< your access token >>",
    "token_type": "bearer",
    "expires_in": 43199,
    "scope": "uaa.resource",
    "jti": "8d00c157058949daab714a44c04c416b"
}
```

12.1.2 Capabilities API

See the list of document fields and enrichment data for each document type you can process with Document Information Extraction.

> → Tip
>
> - See Supported Document Types and File Formats [page 91].
> - See also Supported Languages and Countries/Regions [page 93].

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/capabilities`

HTTP Method: *GET*

Request Parameters

No parameters

Response

Response Fields

JSON Field	Description
documentTypes	List of document types you can submit
enrichment	List of enrichment data you can match
extraction	List of `headerFields`, `lineItemFields`, `enrichment` and `documentTypes`

JSON Field	Description
headerFields	List of header fields you can extract
lineItemFields	List of line items you can extract

The response is given as a status (200, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```json
{
    "extraction":{
        "headerFields":[
            {
                "name":"documentNumber",
                "type":"string",
                "category":"document",
                "supportedDocumentTypes":[
                    "invoice",
                    "paymentAdvice",
                    "purchaseOrder"
                ]
            },
            {
                "name":"taxId",
                "type":"string",
                "category":"amounts",
                "supportedDocumentTypes":[
                    "invoice",
                    "purchaseOrder"
                ]
            },
            {
                "name":"taxName",
                "type":"string",
                "category":"amounts",
                "supportedDocumentTypes":[
                    "invoice"
                ]
            },
            {
                "name":"purchaseOrderNumber",
                "type":"string",
                "category":"details",
                "supportedDocumentTypes":[
                    "invoice"
                ]
            },
            {
                "name":"shippingAmount",
                "type":"number",
                "category":"amounts",
                "supportedDocumentTypes":[
                    "invoice"
                ]
            },
            "..."
        ],
        "lineItemFields":[
            {
                "name":"description",
                "type":"string",
```

```json
            "category":"details",
            "supportedDocumentTypes":[
               "invoice",
               "purchaseOrder"
            ]
         },
         {
            "name":"netAmount",
            "type":"number",
            "category":"amounts",
            "supportedDocumentTypes":[
               "invoice",
               "paymentAdvice",
               "purchaseOrder"
            ]
         },
         {
            "name":"quantity",
            "type":"number",
            "category":"details",
            "supportedDocumentTypes":[
               "invoice",
               "purchaseOrder"
            ]
         },
         {
            "name":"unitPrice",
            "type":"number",
            "category":"details",
            "supportedDocumentTypes":[
               "invoice",
               "purchaseOrder"
            ]
         },
         {
            "name":"materialNumber",
            "type":"string",
            "category":"details",
            "supportedDocumentTypes":[
               "invoice"
            ]
         },
         "..."
      ]
   },
   "enrichment":{
      "employee":{
         "dataTypes":[
            "employee"
         ]
      },
      "sender":{
         "dataTypes":[
            "businessEntity"
         ]
      },
      "receiver":{
         "dataTypes":[
            "businessEntity"
         ]
      }
   },
   "documentTypes":[
      "invoice",
      "paymentAdvice",
      "purchaseOrder",
      "businessCard"
   ]
```

```
    }
```

12.1.3 Client API

Document Information Extraction requires a client to be called. A `default` client is created following tenant provisioning, enabling you to use the service immediately.

The Client API consists of the following endpoints:

-
-
-

12.1.3.1 Create Client

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/clients`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
payload	Yes	JSON Object	*body*	List of clients containing `clientId` and `clientName`

Request Example

Single client:

```
{
    "value":[
        {
            "clientId":"c_00",
            "clientName":"client 00"
        }
    ]
}
```

Multiple clients:

```
{
```

```
    "value":[
        {
            "clientId":"c_00",
            "clientName":"tyler"
        },
        {
            "clientId":"c_01",
            "clientName":"jlaix"
        }
    ]
}
```

Response

Response Fields

JSON Field	Description
inserted	Number of inserted entries
modified	Number of modified entries

The response is given as a status (201, 400, 401, 429, or 500) and JSON file. See Common Status and Error Codes [page 235] and Technical Constraints [page 288].

Response Example

201 "Created"

```
{
    "inserted":1,
    "modified":2
}
```

12.1.3.2 Get Client

Retrieve all the client names and IDs.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/clients`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientIdStar tsWith	No	String	*query*	Filters the list of clients by the characters the clientId starts with. For example: c
limit	Yes	Integer	*query*	Number of clients to process. For example: 10. See Technical Constraints [page 288]
offset	No	Integer	*query*	Index of the first client to be retrieved. For example: 10

Response

Response Fields

JSON Field	Description
id	Tenant ID
payload	List of all clients, including their zoneId, clientId, and clientName

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
  "id": "1234",
  "payload": [
    {
      "clientId": "c_00",
      "clientName": "client 00"
    },
    {
      "clientId": "c_01",
      "clientName": "client 01"
    },
    {
      "clientId": "c_02",
      "clientName": "client 02"
    },
    {
      "clientId": "c_03",
      "clientName": "client 03"
    },
    {
      "clientId": "c_04",
      "clientName": "client 04"
    }
  ]
}
```

12.1.3.3 Delete Client

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/clients`

HTTP Method: *DELETE*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
payload	Yes	JSON Object	*body*	List of client IDs

Response

Response Fields

JSON Field	Description
message	Status message with information about the request

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
  "message": "Successfully deleted 1 client(s)."
}
```

12.1.4 Identifier API (Deprecated)

Create, list, and delete identifiers for client mappings.

> ⚠ **Caution**
>
> This API and its endpoints have been deprecated and are scheduled for decommissioning in June 2025. Please use the Schema API [page 193] to process `paymentAdvice` documents in Excel format. See also Add Fields to Schema Version [page 208].

The Identifier API consists of the following endpoints:

-
-
-

12.1.4.1 Create Identifier (Deprecated)

Create identifiers for client mappings.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/identifier`

HTTP Method: *POST*

Request Parameters

> ⓘ Note
>
> In single POST calls, you can create aliases for only one `documentType` and `fileType`.

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. For example: c_00.
options	Yes	JSON Object	*formData*	Options for processing the document. See the *Options Payload* table below.

Options Payload

Option	Required	Data Type	Description
documentType	Yes	String	Type of the document submitted. For now, only `paymentAdvice` is supported.
fileType	Yes	String	Type of the file submitted. For now, only Excel is supported.

Option	Required	Data Type	Description
headerFields	Yes	String	List of header fields with the aliases of supported capabilities you want to extract from a specific documentType and fileType. See the list of fields that can be extracted from header fields in Extracted Header Fields [page 291].
language	Yes	String	Language of the aliases
lineItemFields	Yes	String	List of line item fields with the aliases of supported capabilities you want to extract from a specific documentType and fileType. See the list of fields that can be extracted from line items in Extracted Line Items [page 300].

Request Example: Options Payload

```
{
    "documentType":"paymentAdvice",
    "fileType":"Excel",
    "headerFields":[
        {
            "language":"en",
            "capabilities":{
                "documentNumber":[
                    "Payment Number"
                ],
                "documentDate":[
                    "Payment Date"
                ],
                "currencyCode":[
                    "Invoice Currency"
                ],
                "grossAmount":[
                    "Amount in Invoice Currency",
                    "Document currency"
                ]
            }
        },
        {
            "language":"de",
            "capabilities":{
                "documentNumber":[
                    "Beleg-Nr."
                ],
                "documentDate":[
                    "RE-Datum"
                ]
            }
        }
    ],
    "lineItemFields":[
        {
            "language":"en",
```

```
            "capabilities":{
                "documentNumber":[
                    "Invoice Number",
                    "Document Number"
                ],
                "documentDate":[
                    "Invoice Date",
                    "Document Date"
                ],
                "discountAmount":[
                    "Cash disc. amt LC"
                ],
                "netAmount":[
                    "Amount Paid",
                    "Amount in doc. curr."
                ]
            }
        },
        {
            "language":"de",
            "capabilities":{
                "documentNumber":[
                    "Beleg-Nr."
                ],
                "documentDate":[
                    "RE-Datum"
                ],
                "netAmount":[
                    "Gesamt-OP"
                ]
            }
        }
    ]
}
```

Response

The response is given as a status (201, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

12.1.4.2 Get Identifier (Deprecated)

Retrieve all identifiers for client mappings by `fileType`, `documentType`, and `clientId`.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/identifier`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`clientId`	Yes	String	*query*	The ID of the client. For example: c_00.
`documentType`	Yes	String	*query*	Type of the document submitted. For now, only `paymentAdvice` is supported.
`fileType`	Yes	String	*query*	Type of the file submitted. For now, only Excel is supported.

Response

Response Fields

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

12.1.4.3 Delete Identifier (Deprecated)

Delete identifiers for client mappings.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/identifier`

HTTP Method: *DELETE*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`clientId`	Yes	String	*query*	The ID of the client. For example: c_00.
`documentType`	No	String	*query*	Type of the document submitted. For now, only `paymentAdvice` is supported.
`fileType`	No	String	*query*	Type of the file submitted. For now, only Excel is supported.

> ⓘ Note
>
> If you want to delete aliases for a specific `documentType` and `fileType`, all parameter fields are required. If the `documentType` and `fileType` are not provided, all aliases are deleted.

Response

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

12.1.5 Configuration API

Create, update, list, and delete configurations on tenant scope by default, or optionally, on instance or client scope.

The Configuration API consists of the following endpoints:

- Create Configuration [page 122]
- Get Configuration [page 127]
- Get Configuration with Key [page 129]
- Delete Configuration [page 131]

Related Information

Configuration Keys [page 124]

12.1.5.1 Create Configuration

Create or update configurations according to the given payload.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/configuration`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	No	String	*query*	The ID of the client you want to set the configuration for. For example: c_00. This parameter is only used for `client` scope configurations.
payload	Yes	JSON Object	*body*	List of configuration key-value pairs. For more information, see Configuration Keys [page 124].
scope	No	String	*query*	Choose the scope of the configuration among the following values: • `client` • `instance` • `tenant` → Tip If you leave this parameter empty, the `tenant` scope is used.
tenantId	No	String	*query*	The ID of the tenant you want to set the configuration for. → Tip If you leave this parameter empty, the `tenantId` sending the request is used.

Response

Response Fields

JSON Field	Description
inserted	Number of inserted entries
modified	Number of modified entries

The response is given as a status (201, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

201 "Success"

```
{
```

```
    "inserted":1,
    "modified":0
}
```

12.1.5.1.1 Configuration Keys

Explore the available configuration keys for the Document Information Extraction service.

Configuration Key	Default Value	Possible Values	Scope	Description	Create Configuration [page 122] Request Payload Example

`activateDocumentNotifications`	**false**	**true**, or **false**	• `client`
			• `instance`
			• `tenant`
				Use this configuration key to enable or disable the Notifications [page 236] functionality. Set `activateDocumentNotifications` to **true** to get notifications about the status of your processed documents.	`{`
					
					`"value":`
					`{`
					
					`"activateDocumentNotifications":`
					`"true"`
					`    }`
					`}`

`clientSegregation`	**false**	**true**, or **false**	• `instance`
			• `tenant`
				Use this configuration key to restrict user access to specified clients. See also Add Document [page 251].	`{`
					
					`"value":`
					`{`
					
					`"clientSegregation":"true"`
					`    }`
					`}`

`coordinateFormat`	**default**	**default**, **absolute**, or **normalized**	• `instance`
			• `tenant`
				Use this configuration key to choose the format of the bounding box coordinates in the extraction results.	`{`
					
					`"value":`
					`{`
					
					`"coordinateFormat":"normalized"`
					`    }`
					`}`

Configuration Key	Default Value	Possible Values	Scope	Description	Create Configuration [page 122] Request Payload Example
`dataFeedbackCollection`	**false**	**true**, or **false**	• `client` • `instance` • `tenant`	Use this configuration key to make use of the data feedback collection feature. See also Confirm Document [page 165]. If set to **false**, all documents already uploaded to the service for retraining by this tenant (or instance) are deleted. And all documents uploaded from that moment onwards are no longer used to retrain the service's machine learning models. See also Delete Configuration [page 131]. → Remember As Document Information Extraction learns from data, enabling data feedback collection may help the service to become more accurate in extracting information from your documents. On the contrary, deletion of data may result in extraction results becoming less accurate. Deletion of data is irreversible.	`{ "value": { "dataFeedbackCollection":"true" } }`
`documentRetentionTimeDays`	**7** days	**1** to **30** days	• `client` • `instance` • `tenant`	Use this configuration key to set the retention period for inference documents uploaded to the service.	`{ "value": { "documentRetentionTimeDays":"10" } }`

Document Information Extraction
Development

PUBLIC **125**

Configuration Key	Default Value	Possible Values	Scope	Description	
enrichment Confidence Threshold	low	low, medium, or high	• client • instance • tenant	Use this configuration key to adjust the similarity confidence threshold for the enrichment. The **low** value results in more matches with higher possiblity of false-positve matches. The **high** value returns only very confident matches and has lower tolerance for differences between document content and master data. The **medium** value is a balanced adjustment. This configuration can alter the behavior of the enrichment. If you don't get good enrichment results, it's recommended to test the different values for this configuration. Use a lower value if you want to get more matches, or if the expected master data doesn't match the document. Use a higher value if you get incorrect or unexpected matches.	```{ "value": { "enrichm entConfi denceThr eshold": "medium" } }```
manualData Activation	false	true, or false	tenant	Use this configuration key to set data activation to manual, instead of using the default automatic refresh of enrichment data, that takes place every 4 hours. See also Create Data Activation [page 187] and Get Data Activation Details [page 188].	```{ "value": { "manualD ataActiv ation":" true" } }```

Configuration Key	Default Value	Possible Values	Scope	Description	Request Payload Example
performPII Check	**true**	**true**, or **false**	tenant	This is a subconfiguration of the `dataFeedbackCollection` configuration key. To use this subconfiguration, set the `dataFeedbackCollection` configuration key to **true**. The `performPIICheck` subconfiguration is set to **true** by default. If set to true, the service automatically scans documents for Personally Identifiable Information (PII) and excludes any document with PII from being used for improving the service. If you set `performPIICheck` to **false**, all documents may be used for improving the service. See also Confirm Document [page 165].	```{ "value": { "perform PIICheck ":"false " } }```

> ⓘ Note
>
> Before setting the `dataFeedbackCollection` configuration key to *true*, and the `performPIICheck` subconfiguration to *false*, review the subsection *Deletion of Personal Data* in Data Protection and Privacy [page 302].

> ⚠ Restriction
>
> The `documentRetentionTimeDays` and `dataFeedbackCollection` configuration keys, and the `performPIICheck` subconfiguration are only available on production environments for enterprise accounts. These keys are not available for trial account users.

12.1.5.2 Get Configuration

Retrieve all configurations already created for the requested scope.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/configuration`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`clientId`	No	String	*query*	The ID of the client you want to get the configuration for. For example: c_00. This parameter is only used for client `scope` configurations.
`scope`	No	String	*query*	Choose the scope of the configuration among the following values: • `active` (all configurations already created, including the ones on `client`, `instance`, and `tenant` scope) • `client` • `instance` • `tenant` → Tip If you leave this parameter empty, the `active` scope is used.
`tenantId`	No	String	*query*	The ID of the tenant you want to get the configuration for. → Tip If you leave this parameter empty, the `tenantId` sending the request is used.

Response

Response Fields

JSON Field	Description
`results`	List containing information of all configurations

The response is given as a status (200, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "results":{
        "documentRetentionTimeDays": "10",
        "manualDataActivation": "true",
        "dataFeedbackCollection": "true",
        "performPIICheck": "true"
    }
}
```

12.1.5.3 Get Configuration with Key

Retrieve all configurations already created for a given key for the requested scope.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/configuration/<key>`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	No	String	*query*	The ID of the client you want to get the configuration for. For example: c_00. This parameter is only used for client `scope` configurations.
key	Yes	String	*query*	One of the available Configuration Keys [page 124].

Parameter	Required	Data Type	Parameter Type	Description
scope	No	String	*query*	Choose the scope of the configuration among the following values: • `active` (all configurations already created, including the ones on `client`, `instance`, and `tenant` scope) • `client` • `instance` • `tenant` → **Tip** If you leave this parameter empty, the `active` scope is used.
tenantId	No	String	*path*	The ID of the tenant you want to get the configuration for. → **Tip** If you leave this parameter empty, the `tenantId` sending the request is used.

Response

Response Fields

JSON Field	Description
results	List containing information of all configurations

The response is given as a status (200, 400, 401, 404, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "results":{
        "documentRetentionTimeDays": "10"
    }
}
```

```
{
    "results":{
```

```
        "manualDataActivation": "true"
    }
}
```

```
{
    "results":{
        "dataFeedbackCollection": "true"
    }
}
```

```
{
    "results":{
        "performPIICheck": "true"
    }
}
```

12.1.5.4 Delete Configuration

Delete configurations according to the given payload.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/configuration`

HTTP Method: *DELETE*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	No	String	*query*	The ID of the client you want to delete the configuration for. For example: c_00. This parameter is only used for client scope configurations.

Parameter	Required	Data Type	Parameter Type	Description
`payload`	Yes	JSON Object	*body*	List of configuration keys. All configurations are deleted if payload is empty. Possible configuration key and subconfiguration values:

- `activateDocumentNotifications`
- `clientSegregation`
- `coordinateFormat`
- `dataFeedbackCollection`

> ⓘ **Note**
>
> After sending the *DELETE* request using the `dataFeedbackCollection` configuration key, all documents already uploaded to the service for retraining by this tenant (or service instance) are deleted. And all documents uploaded from that moment onwards are no longer used to retrain the service's machine learning models. See also Configuration Keys [page 124] (if parameter is set to *false*).

- `documentRetentionTimeDays`

> ⓘ **Note**
>
> After sending the *DELETE* request using the `documentRetentionTimeDays` configuration key, the default retention period of 7 days is used again.

- `enrichmentConfidenceThreshold`
- `manualDataActivation`
- `performPIICheck`

Parameter	Required	Data Type	Parameter Type	Description
scope	No	String	*query*	Choose the scope of the configuration among the following values: • `client` • `instance` • `tenant` → Tip If you leave this parameter empty, the `tenant` scope is used.
tenantId	No	String	*query*	The ID of the tenant you want to delete the configuration for. → Tip If you leave this parameter empty, the `tenantId` sending the request is used.

Request Examples

```
{
    "value":[
        "documentRetentionTimeDays"
    ]
}
```

```
{
    "value":[
        "manualDataActivation"
    ]
}
```

```
{
    "value":[
        "dataFeedbackCollection"
    ]
}
```

```
{
    "value":[
        "performPIICheck"
    ]
}
```

```
{
    "value":[
        "documentRetentionTimeDays",
        "manualDataActivation",
        "dataFeedbackCollection",
        "performPIICheck"
```

```
    ]
}
```

Response

Response Fields

JSON Field	Description
deleted	Total number of configurations deleted with this request

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
   "deleted": 1
}
```

12.1.6 Document API

The core functionality of Document Information Extraction is extracting structured information from documents automatically using machine learning. The Document API provides endpoints to upload documents for processing and also to get the results.

The Document API consists of the following endpoints:

Document Information Extraction
Development

12.1.6.1 Upload Document

Upload a document file to the service to get the extraction results from header fields and line items in JSON format.

> → Tip
>
> - See Supported Document Types and File Formats [page 91].
> - See also Supported Languages and Countries/Regions [page 93].

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/document/jobs`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
file	Yes	File	*formData*	Document file you want to process. See Supported Document Types and File Formats [page 91].
options	Yes	JSON Object	*formData*	Options for processing the document. See the *Options Payload* table below.

Options Payload

Option	Required	Data Type	Description
candidateTemplateIds	No	String	IDs of templates from which the service automatically detects the appropriate `templateId`.
clientId	Yes	String	The ID of the client. For example: c_00
customLabel	No	String	The label you want to use for the document. If used you can query the corresponding document ID using the Post Catalog [page 142] endpoint.
documentType	No	String	The type of the document you uploaded. For example: `invoice` or `paymentAdvice`

Option	Required	Data Type	Description
enrichment	No	String	See Enrichment Parameter [page 141].
headerFields	No (Yes if payload doesn't include schemaId or schemaName)	String	Comma-separated list of header fields you want to extract. When you include schemaId in the payload, don't include a list of headerFields. See the list of fields that can be extracted from header fields in Extracted Header Fields [page 291].
lineItemFields	No (Yes if payload doesn't include schemaId or schemaName)	String	Comma-separated list of line item fields you want to extract. When you include schemaId in the payload, don't include a list of lineItemFields. See the list of fields that can be extracted from line items in Extracted Line Items [page 300].
receivedDate	No	String	The date when the document was received. For example: 2020-02-17

Option	Required	Data Type	Description
templateId	No	String	The ID of the template to be used for this document.

<div style="margin-left: 50%;">

To detect `templateId` automatically, use the value `"detect"` instead of the ID string. You can also optionally use the `candidateTemplateIds` option to restrict detection to specified templates.

> ⚠ **Caution**
>
> `schemaId` isn't always a required option. However, if your payload includes `templateId`, it must also include `schemaId` or `schemaName`. In such cases, don't include `headerFields` or `lineItemFields` in the payload to avoid conflicts.

</div>

Option	Required	Data Type	Description
schemaId	No (Yes if payload includes templateId or doesn't include a list of headerFields and / or a list of lineItemFields, or schemaName)	String	The ID of the schema to be used for this document. To use one of the preconfigured SAP schemas, consider the following schema IDs and document types: • SAP_OCROnly_schema: "schemaId":"09e6c9e4-d7b0-414f-bd85-cfee6fbb2add" for custom documents • SAP_invoice_schema: "schemaId":"cf8cc8a9-1eee-42d9-9a3e-507a61baac23" for invoice documents • SAP_purchaseOrder_schema: "schemaId":"fbab052e-6f9b-4a5f-b42f-29a8162eb1bf" for purchaseOrder documents • SAP_paymentAdvice_schema: "schemaId":"b7fdcfac-7853-42bb-89d2-ede2ba1ce803" for paymentAdvice documents
schemaName	No (Yes if payload includes templateId or doesn't include a list of headerFields and / or a list of lineItemFields, or schemaId)	String	The name of the schema to be used for this document. To use one of the preconfigured SAP schemas, consider the schema names and document types listed above in the description for schemaId.

Option	Required	Data Type	Description
schemaVersion	No (Yes if payload includes schemaId or schemaName, and you don't want to use the schemaVersion default version 1)	String	The version number of the schema you want to use for this document. In the payload, schemaId or schemaName, and schemaVersion must be provided. If schemaVersion isn't provided, default version 1 is used.

Request Example: Options Payload for Autodetecting templateID

```
{
   "clientId":"c_00",
   "documentType":"invoice",
   "receivedDate":"2020-02-17",
   "schemaId":"10c10bd2-082b-47c8-851d-e58827828637",
   "templateId":"detect",
   "candidateTemplateIds":[
      "0ebcd5c4-7843-4e6e-867a-1e5c997e4e4c",
      "98ee6ff3-30bf-4e22-8579-0f0bde462c53",
      "d6f62ef3-551a-454d-bfa4-fc334af30bf2"
   ],
   "enrichment":{

   }
}
```

Request Example: Options Payload with Template

```
{
   "clientId":"c_00",
   "documentType":"invoice",
   "receivedDate":"2020-02-17",
   "schemaId":"10c10bd2-082b-47c8-851d-e58827828637",
   "templateId":"0ebcd5c4-7843-4e6e-867a-1e5c997e4e4c",
   "enrichment":{

   }
}
```

Request Example: Options Payload without Template

```
{
   "extraction":{
      "headerFields":[
         "documentNumber",
         "taxId",
         "purchaseOrderNumber",
         "shippingAmount",
         "netAmount",
         "senderAddress",
         "senderName",
         "grossAmount",
         "currencyCode",
         "receiverContact",
         "documentDate",
         "taxAmount",
```

```
        "taxRate",
        "receiverName",
        "receiverAddress"
    ],
    "lineItemFields":[
        "description",
        "netAmount",
        "quantity",
        "unitPrice",
        "materialNumber"
    ]
  },
  "clientId": "c_00",
  "documentType": "invoice",
  "receivedDate": "2020-02-17",
  "enrichment": {}
}
```

Request Example: Options Payload with the `schemaName` **of** `SAP_OCROnly_schema`

```
{
    "schemaName":"SAP_OCROnly_schema",
    "clientId":"c_10",
    "documentType":"custom"
}
```

Request Example: Options Payload with the `schemaId` **of** `SAP_OCROnly_schema`

```
{
    "schemaId":"09e6c9e4-d7b0-414f-bd85-cfee6fbb2add",
    "clientId":"c_10",
    "documentType":"custom"
}
```

Response

Response Fields

JSON Field	Description
id	Request ID
processedTime	Timestamp in RFC format
status	Status of the request. Possible values: "PENDING", "DONE", or "FAILED"

The response is given as a status (201, 400, 401, 415, 429, 500, or 503). See Common Status and Error Codes [page 235].

Response Example

201 "Created"

```
{
   "id":"484b6e1c-501c-4a07-85cb-84554656a175",
   "status":"PENDING",
   "processedTime":"2020-03-26T17:00:00.000000+00:00"
}
```

12.1.6.1.1 Enrichment Parameter

The enrichment parameter can be used to retrieve a matching of enrichment data to extracted header fields. See Create Enrichment Data [page 175]. The property should be a JSON object which can contain properties, as listed in the table below, depending on the enrichment data you want to match.

Example

```
"enrichment":{
    "sender":{
        "top":5,
        "type":"businessEntity",
        "subtype":"supplier"
    },
    "employee":{
        "type":"employee"
    },
    "product":{
        "type":"product"
    }
}
```

Property	Required	Description
employee	No	To match the *receiverContact* extracted header field to enrichment data, the `employee` property should be present in enrichment.
product	No	To match the product line items found on the document to enrichment data, the `product` property should be present in enrichment.
receiver	No	To match the extracted visual information from the document to the receiver enrichment data, the `receiver` property should be present in enrichment.
sender	No	To match the extracted visual information from the document to the sender enrichment data, the `sender` property should be present in enrichment.
type	Yes	The type of enrichment data entities used for matching. Available values: `businessEntity`, `employee`, and `product`. See Entities [page 178] for details about the available enrichment data entity types.
subtype	No	The subtype of enrichment data entities used for matching with type **businessEntity**. Available values: `supplier`, `customer`, and `companyCode`.

Property	Required	Description
`top`	No	The `top` property specifies the maximum number of matched enrichment data entities returned.

> ⓘ **Note**
>
> If the top property is not defined, the default value is 1. The maximum possible value of the property is 50. If you enter a value higher than 50, you will get an error message with the maximum possible value.

> ⓘ **Note**
>
> The following properties are optional, but, in case you want to match enrichment data, at least one of them is required:
>
> - `sender`
> - `receiver`
> - `employee`
> - `product`

Related Information

12.1.6.2 Post Catalog

Post a search or filter request to get the current status of document processing jobs. Returns a list with all document processing jobs in a JSON file.

Optionally, the jobs can be filtered based on the client ID and a filter query. You have the following catalog options:

- Filtering using the `filter` and `likeFilter` parameters
- Ordering using the `order` parameter
- Pagination using the `limit` and `offset` parameters

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/document/catalog`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
options	Yes	JSON Object	*formData*	Catalog options used when searching for documents. See the *Options Payload* table below.

Options Payload

Option	Required	Data Type	Description
clientId	No	String	The ID of the client used while submitting the document. For example: "c_00".
filter	No	String	Filter query for retrieving documents. The filter query needs to be an expression. The expression should follow the format: "fieldName op value <AND/OR> fieldName op value". Supported fields: documentType, created, schemaId, status, customLabel, reviewStatus, or tenantId. Possible operators (op): depend on the field. For example: "status eq done".
likeFilter	No	String	Filter query for retrieving documents that uses the LIKE operator. The expression should follow the format "fieldName like value". Supported field: fileName. For example: "fileName like \"test receipt\"".
limit	No	Integer	Number of documents to retrieve (maximum allowed value: 50). For example: 10.
offset	No	Integer	Index of the first document to be retrieved. For example: 20.
order	No	String	Order criteria for the retrieved documents. Possible values: created, fileName, documentType, or status. For example: "created asc" (sorts by creation date in ascending order).

Request Example: Options Payload

```
{
   "clientId":"c_00",
   "limit":10,
   "offset":2,
   "order":"created desc",
   "likeFilter":"fileName like \"test receipt\"",
   "filter":"status eq failed or documentType eq invoice"
}
```

Response

Response Fields

JSON Field	Description
results	List containing all document processing jobs
totalDocumentCount	Total number of document processing jobs returned by the request options
usedOptions	Options used in the filtering and/or ordering of document processing jobs

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
   "results":[
      [
         {
            "status":"DONE",
            "id":"c4f25368-d3e6-43f7-a0b4-55adf7f54e95",
            "fileName":"test receipt_invoice1.pdf",
            "documentType":"invoice",
            "created":"2020-03-26 17:00:00.000000+00:00",
            "clientId":"c_00",
            "finished":"2020-03-26 17:01:30.000000+00:00"
         },
         {
            "status":"PENDING",
            "id":"50199d80-c742-453b-830d-8e6ce14568e2",
            "fileName":"test receipt invoice2.pdf",
            "documentType":"invoice",
            "created":"2020-03-26 18:00:00.000000+00:00",
            "clientId":"c_00"
         },
         {
            "status":"FAILED",
            "id":"50199d80-c742-453b-830d-8e6ce14568e2",
            "fileName":"test receipt pa.pdf",
            "documentType":"paymentAdvice",
            "created":"2020-03-26 19:00:00.000000+00:00",
            "clientId":"c_00",
            "finished":"2020-03-26 19:01:30.000000+00:00"
         }
      ]
```

```
    ],
    "usedOptions":{
        "clientId":"c_00",
        "limit":10,
        "offset":2,
        "order":"created desc",
        "likeFilter":"fileName like \"test receipt\"",
        "filter":"status eq failed or documentType eq invoice"
    },
    "totalDocumentCount":5
}
```

12.1.6.3 List Documents

Get a list of up to 200 documents in a JSON file.

> → Tip
>
> Use the endpoint Post Catalog [page 142] to page through lists of more than 200 documents in a JSON file.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/document/jobs`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	No	String	*query*	The ID of the client. For example: c_00

Response

Response Fields

JSON Field	Description
clientId	ID of the client this document was processed for
created	Time when the document was submitted for processing
finished	Time when the document status changed to "DONE", or "FAILED"
id	Document ID

JSON Field	Description
results	List containing information of all processed documents
status	Processing status of the document. Possible values: "PENDING", "DONE", "CONFIRMED", or "FAILED"

The response is given as a status (200, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "results":[
        [
            {
                "id":"c4f25368-d3e6-43f7-a0b4-55adf7f54e95",
                "clientId":"c1",
                "created":"2020-05-08T10:39:59.916359+00:00",
                "finished":"2020-05-08T10:40:50.467719+00:00",
                "status":"DONE"
            },
            {
                "id":"50199d80-c742-453b-830d-8e6ce14568e2",
                "clientId":"c1",
                "created":"2020-05-12T08:30:04.718730+00:00",
                "status":"PENDING"
            },
            {
                "id":"47299d80-c742-453b-830d-8e6ce14568e2",
                "clientId":"c1",
                "created":"2020-05-12T08:23:06.938779+00:00",
                "finished":"2020-05-12T08:23:21.765680+00:00",
                "status":"FAILED"
            }
        ]
    ]
}
```

12.1.6.4 Get Result

The Document Information Extraction service takes document files as input and returns a JSON file that contains the information that has been extracted from the header fields and line items of the specified document. See Supported Document Types and File Formats [page 91].

> → Remember
>
> Document Information Extraction typically provides extraction results for an average document in about 30 seconds.
>
> However, processing can take longer if the task involved is more complex – for example, if the documents processed are large.
>
> Before you use the service for important or time-sensitive tasks, we strongly recommend running mass tests to assess the performance of the service and make sure it meets your requirements.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/document/jobs/<id>`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`extractedValues`	No	Boolean	*query*	Set to *true* to get the extracted values. Set to false to get the ground truth values, if available. If ground truth values are not available, extracted values are returned in any case. The default value for this parameter is *false*.
`id`	Yes	String	*path*	The ID returned by the Upload Document [page 135] endpoint. For example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff.
`returnNullValues`	No	Boolean	*query*	Set to *true* to get all requested fields in the document results, even if they could not be extracted. For fields that could not be extracted, for example, because they are not available in the document or because the service was not able to identify the field, the value is **null**.

If no `value` can be extracted, both `value` and `rawValue` are **null**. If a `value` is extracted, it can occur that the corresponding `rawValue` is displayed as an empty string.

Additionally, the prediction confidence score is **null** and the x-coordinate, y-coordinate, width, and height are set to **0**. Per default, this parameter is set to *false* and fields not extracted are not returned. |

Response

Response Fields

JSON Field	Description
attributes	Dictionary containing the `method` of the matched enrichment data record. Or dictionary containing the `symbology` of the extracted barcode header field.
bocrVersion	The version number of the Optical Character Recognition (OCR) service.
category	Category of the field. For example: document or receiver.
clientId	Identify the client that submitted the extraction request using the Upload Document [page 135] endpoint.
confidence	Prediction confidence score for a field or enrichment data. The possible values are between 0.0 and 1.0.
coordinates	Bounding box coordinates for this field (not present if value is **null**).
country	Country/Region code of the document submitted.
created	Time when the document was submitted for processing.
dataForRetrainingStatus	Retraining status. Possible values: "notUsedForTraining", "rejectedDueToPII", "inProcess", "acceptedForTraining", or "usedForTraining".
documentType	Type of the document submitted.
doxVersion	The version number of the Document Information Extraction service.
employee	Employee enrichment data. For example: employee name.
enrichment	Dictionary containing enrichment data.
extraction	Dictionary containing all the extracted header fields and line items.
fileName	Full name of the document submitted.
fileType	File format of the document submitted. For example: PDF, PNG, JPEG.
finished	Time when the document status changed to "DONE", or "FAILED".
group	Group this field belongs to.
headerFields	Dictionary containing all extracted header fields.
height	Page height of the document
id	Document or enrichment data ID.
label	User-friendly names for header fields and line items. See Add Fields to Schema Version [page 208].
languageCodes	Array containing strings of language codes. For example: "en" for English and "de" for German.
lineItems	Dictionary containing all extracted line items.
method	Match strategy for each matched enrichment data record. Possible values: "exactTaxId", "exactBankAccount", "exactMaterialNumber", or "similarity".

JSON Field	Description
`model`	The model used to extract information from the specified field. Possible values: "ai" or "template". "ai" denotes the machine learning models of the Document Information Extraction service.
`name`	Name of the field.
`page`	Page number of the document where the field was found (not present if value is **null**).
`pageCount`	Total number of pages a document contains. For example: 2.
`rawValue`	Value extracted for this field by the Document Information Extraction service as displayed in the document.
`schemaId`	The ID of the schema used when you uploaded the document.
`schemaName`	The name of the schema used when you uploaded the document.
`schemaVersion`	The version number of the schema used when you uploaded the document.
`sender`	Sender enrichment data. For example: sender name and sender address.
`status`	Processing status of the document. Possible values: "PENDING", "DONE", or "FAILED".
`symbology`	Type of the extracted barcode. For example: QR.
`templateId`	The ID of the template associated with the document
`type`	Data type of the extracted hearder fields and line items.
`value`	Value extracted for this field by the Document Information Extraction service in standardized format.
`values`	Dictionary containing all matched enrichment data records.
`variant`	See Data Variants [page 181].
`width`	Page width of the document

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success" with `SAP_OCROnly_schema` ("schemaId":"09e6c9e4-d7b0-414f-bd85-cfee6fbb2add")

```
{
    "status":"DONE",
    "id":"2acc2040-f956-4178-9cf4-d02f020626a6",
    "fileName":"sample-power_of_attorney-3.pdf",
    "documentType":"custom",
    "created":"2022-10-04T07:46:03.412498+00:00",
    "finished":"2022-10-04T07:46:56.834313+00:00",
    "clientId":"c_00",
    "languageCodes":[
        "en"
    ],
    "pageCount":1,
    "schemaId":"09e6c9e4-d7b0-414f-bd85-cfee6fbb2add",
    "schemaName":"SAP_OCROnly_schema",
    "country":null,
    "extraction":{
        "headerFields":[
```

```
        ],
        "lineItems":[

        ]
    },
    "bocrVersion":"2.7.1",
    "doxVersion":"local",
    "fileType":"pdf",
    "dataForRetrainingStatus":"notUsedForTraining"
}
```

Response Example

200 "Success" without `schemaId`

```
{
    "status":"DONE",
    "id":"a712375f-0b6d-4550-83fb-2271a2301aad",
    "fileName":"demo_taxid.pdf",
    "documentType":"invoice",
    "created":"2022-04-27T09:46:20.090953+00:00",
    "finished":"2022-04-27T09:46:45.151654+00:00",
    "clientId":"c_00",
    "languageCodes":[
        "xx"
    ],
    "pageCount":1,
    "country":"MX",
    "extraction":{
        "headerFields":[
            {
                "name":"receiverContact",
                "category":"receiver",
                "value":"FELISHIA ICE SOUTH",
                "rawValue":"FELISHIA ICE SOUTH",
                "type":"string",
                "page":1,
                "confidence":0.9431540966033936,
                "coordinates":{
                    "x":0.0792156862745098,
                    "y":0.19575757575757577,
                    "w":0.13568627450980392,
                    "h":0.009090909090909066
                },
                "model":"ai",
                "label":"receiverContact"
            },
            {
                "name":"receiverName",
                "category":"receiver",
                "value":"JOERNS ICE HEALTH (EST)",
                "rawValue":"JOERNS ICE HEALTH (EST)",
                "type":"string",
                "page":1,
                "confidence":0.8918973803520203,
                "coordinates":{
                    "x":0.05215686274509804,
                    "y":0.45696969696969697,
                    "w":0.18823529411764706,
                    "h":0.011818181818181839
                },
                "model":"ai",
                "label":"receiverName"
            },
            {
                "name":"shippingAmount",
                "category":"amounts",
```

```
            "value":0.0,
            "rawValue":"0.00",
            "type":"number",
            "page":1,
            "confidence":0.9837643504142761,
            "coordinates":{
               "x":0.907843137254902,
               "y":0.7975757575757576,
               "w":0.03215686274509799,
               "h":0.009393939393939399
            },
            "model":"ai",
            "label":"shippingAmount"
         },
         {
            "name":"taxAmount",
            "category":"amounts",
            "value":7.07,
            "rawValue":"7.07",
            "type":"number",
            "page":1,
            "confidence":0.9896121621131897,
            "coordinates":{
               "x":0.9078431129455566,
               "y":0.8166666626930237,
               "w":0.0313725471496582,
               "h":0.008787870407104492
            },
            "model":"ai",
            "group":1,
            "label":"taxAmount"
         },
         {
            "name":"senderAddress",
            "category":"sender",
            "value":"12345 NETWORK PLACE CHICAGO, IL 60000-1234",
            "rawValue":"12345 NETWORK PLACE CHICAGO, IL 60000-1234",
            "type":"string",
            "page":1,
            "confidence":0.6106114352383017,
            "coordinates":{
               "x":0.3184313725490196,
               "y":0.1087878787878788,
               "w":0.15725490196078434,
               "h":0.022727272727272707
            },
            "model":"ai",
            "label":"senderAddress"
         },
         {
            "name":"receiverAddress",
            "category":"receiver",
            "value":"12345 DEARBORN STREET CHATSWORTH, CA 12345 UNITED STATES",
            "rawValue":"12345 DEARBORN STREET CHATSWORTH, CA 12345 UNITED
STATES",
            "type":"string",
            "page":1,
            "confidence":0.5784785588744978,
            "coordinates":{
               "x":0.07882352941176471,
               "y":0.21363636363636362,
               "w":0.21607843137254903,
               "h":0.04545454545454547
            },
            "model":"ai",
            "label":"receiverAddress"
         },
         {
```

```
            "name":"senderName",
            "category":"sender",
            "value":"GLOBAL ICE COMPANY INC.",
            "rawValue":"GLOBAL ICE COMPANY INC.",
            "type":"string",
            "page":1,
            "confidence":0.602843187909389,
            "coordinates":{
               "x":0.343921568627451,
               "y":0.2875757575757576,
               "w":0.2984313725490196,
               "h":0.009393939393939399
            },
            "model":"ai",
            "label":"senderName"
         },
         {
            "name":"taxId",
            "category":"amounts",
            "value":"11-3584699",
            "rawValue":"11-3584699",
            "type":"string",
            "page":1,
            "confidence":0.950018584728241,
            "coordinates":{
               "x":0.5015686274509804,
               "y":0.3409090909090909,
               "w":0.08784313725490189,
               "h":0.009393939393939399
            },
            "model":"ai",
            "group":1,
            "label":"taxId"
         },
         {
            "name":"currencyCode",
            "category":"amounts",
            "value":"USD",
            "rawValue":"",
            "type":"string",
            "page":1,
            "confidence":0.9978113174438477,
            "coordinates":{
               "x":0.0,
               "y":0.0,
               "w":0.0,
               "h":0.0
            },
            "model":"ai",
            "label":"currencyCode"
         },
         {
            "name":"documentNumber",
            "category":"document",
            "value":"112857784",
            "rawValue":"112857784",
            "type":"string",
            "page":1,
            "confidence":0.9963446855545044,
            "coordinates":{
               "x":0.5862745098039216,
               "y":0.08757575757575757,
               "w":0.0811764705882353,
               "h":0.008787878787878789
            },
            "model":"ai",
            "label":"documentNumber"
         },
```

```json
{
    "name":"documentDate",
    "category":"document",
    "value":"2018-06-29",
    "rawValue":"06-29-2018",
    "type":"date",
    "page":1,
    "confidence":0.9906787872314453,
    "coordinates":{
        "x":0.7003921568627451,
        "y":0.08757575757575757,
        "w":0.08470588235294119,
        "h":0.008484848484848484
    },
    "model":"ai",
    "label":"documentDate"
},
{
    "name":"grossAmount",
    "category":"amounts",
    "value":108.13,
    "rawValue":"108.13",
    "type":"number",
    "page":1,
    "confidence":0.9433890581130981,
    "coordinates":{
        "x":0.8913725490196078,
        "y":0.8357575757575758,
        "w":0.05058823529411771,
        "h":0.009696969696969648
    },
    "model":"ai",
    "label":"grossAmount"
},
{
    "name":"netAmount",
    "category":"amounts",
    "value":101.06,
    "rawValue":"101.06",
    "type":"number",
    "page":1,
    "confidence":0.9396025538444519,
    "coordinates":{
        "x":0.8901960784313725,
        "y":0.7778787878787878,
        "w":0.04980392156862745,
        "h":0.010000000000000009
    },
    "model":"ai",
    "label":"netAmount"
},
{
    "name":"purchaseOrderNumber",
    "category":"details",
    "value":"14035740",
    "rawValue":"14035740",
    "type":"string",
    "page":1,
    "confidence":0.7348883748054504,
    "coordinates":{
        "x":0.5905882352941176,
        "y":0.12818181818181817,
        "w":0.0725490196078431,
        "h":0.009090909090909094
    },
    "model":"ai",
    "label":"purchaseOrderNumber"
}
```

```
      ],
      "lineItems":[
         [
            {
               "name":"description",
               "category":"details",
               "value":"PIP Ambi-Dex&#174; 63-331PF Industrial Grade Nitrile
Gloves, Powder-Free, Tex- tured, Blue, L, 100/Box - Tracking#:
1ZX647100300080084",
               "rawValue":"PIP Ambi-Dex&#174; 63-331PF Industrial Grade
Nitrile Gloves, Powder-Free, Tex- tured, Blue, L, 100/Box - Tracking#:
1ZX647100300080084",
               "type":"string",
               "page":1,
               "confidence":0.8756256103515625,
               "coordinates":{
                  "x":0.2988235294117647,
                  "y":0.6545454545454545,
                  "w":0.30980392156862746,
                  "h":0.05363636363636359
               },
               "model":"ai",
               "label":"description"
            },
            {
               "name":"materialNumber",
               "category":"details",
               "value":"B676817",
               "rawValue":"B676817",
               "type":"string",
               "page":1,
               "confidence":0.982785165309906,
               "coordinates":{
                  "x":0.18313725490196078,
                  "y":0.654848484848849,
                  "w":0.06627450980392158,
                  "h":0.009393939393939288
               },
               "model":"ai",
               "label":"materialNumber"
            },
            {
               "name":"netAmount",
               "category":"amounts",
               "value":88.0,
               "rawValue":"88.00",
               "type":"number",
               "page":1,
               "confidence":0.8774160146713257,
               "coordinates":{
                  "x":0.779607843137255,
                  "y":0.6551515151515152,
                  "w":0.04117647058823526,
                  "h":0.009393939393939399
               },
               "model":"ai",
               "label":"netAmount"
            },
            {
               "name":"quantity",
               "category":"details",
               "value":10.0,
               "rawValue":"10",
               "type":"number",
               "page":1,
               "confidence":0.9688798189163208,
               "coordinates":{
                  "x":0.08627450980392157,
```

```
                        "y":0.6551515151515152,
                        "w":0.016470588235294126,
                        "h":0.009090909090909038
                    },
                    "model":"ai",
                    "label":"quantity"
                },
                {
                    "name":"unitPrice",
                    "category":"details",
                    "value":8.8,
                    "rawValue":"8.80",
                    "type":"number",
                    "page":1,
                    "confidence":0.9341872334480286,
                    "coordinates":{
                        "x":0.6862745098039216,
                        "y":0.6551515151515152,
                        "w":0.03176470588235292,
                        "h":0.009393939393939399
                    },
                    "model":"ai",
                    "label":"unitPrice"
                }
            ],
            [
                {
                    "name":"description",
                    "category":"details",
                    "value":"Ergodyne&#174; ProFlex&#174; 812 Standard Utility Glove,
Black, Large, 17174 - Tracking#: 1ZX647100300081467",
                    "rawValue":"Ergodyne&#174; ProFlex&#174; 812 Standard Utility
Glove, Black, Large, 17174 - Tracking#: 1ZX647100300081467",
                    "type":"string",
                    "page":1,
                    "confidence":0.7070900797843933,
                    "coordinates":{
                        "x":0.296078431372549,
                        "y":0.7233333333333334,
                        "w":0.323921568627451,
                        "h":0.040303030303030285
                    },
                    "model":"ai",
                    "label":"description"
                },
                {
                    "name":"materialNumber",
                    "category":"details",
                    "value":"B2139393",
                    "rawValue":"B2139393",
                    "type":"string",
                    "page":1,
                    "confidence":0.9847809076309204,
                    "coordinates":{
                        "x":0.17882352941176471,
                        "y":0.7233333333333334,
                        "w":0.07529411764705879,
                        "h":0.009696969696969648
                    },
                    "model":"ai",
                    "label":"materialNumber"
                },
                {
                    "name":"netAmount",
                    "category":"amounts",
                    "value":13.06,
                    "rawValue":"13.06",
                    "type":"number",
```

```
            "page":1,
            "confidence":0.8856437802314758,
            "coordinates":{
                "x":0.7803921568627451,
                "y":0.7233333333333334,
                "w":0.040784313725490184,
                "h":0.010303030303030258
            },
            "model":"ai",
            "label":"netAmount"
        },
        {
            "name":"quantity",
            "category":"details",
            "value":1.0,
            "rawValue":"1",
            "type":"number",
            "page":1,
            "confidence":0.9791963696479797,
            "coordinates":{
                "x":0.08901960784313726,
                "y":0.723939393939394,
                "w":0.010196078431372546,
                "h":0.008484848484848428
            },
            "model":"ai",
            "label":"quantity"
        },
        {
            "name":"unitPrice",
            "category":"details",
            "value":13.06,
            "rawValue":"13.06",
            "type":"number",
            "page":1,
            "confidence":0.8986196517944336,
            "coordinates":{
                "x":0.6772549019607843,
                "y":0.7233333333333334,
                "w":0.04117647058823526,
                "h":0.010303030303030258
            },
            "model":"ai",
            "label":"unitPrice"
        }
    ]
  ]
},
"bocrVersion":null,
"doxVersion":"local",
"fileType":"pdf",
"enrichment":{
    "sender":[
        {
            "id":"demo-match",
            "confidence":0.7157895,
            "values":{
                "name":"GLOBAL ICE COMPANY INC.",
                "bankAccount":"de23672700030136040305",
                "email":"example@sap.com",
                "address1":"12345 NETWORK PLACE CHICAGO, IL 60673-1298 2",
                "countryCode":"US",
                "state":"Illinois",
                "city":"Chicago",
                "postalCode":"60007"
            },
            "attributes":{
                "method":"similarity"
```

```
                }
            }
        ],
        "employee":[

        ],
        "product":[

        ]
    },
    "dataForRetrainingStatus":"notUsedForTraining"
}
```

If the document is processed successfully, Document Information Extraction provides the predictions for the requested fields. The requested fields are those which were requested in . When no value can be detected for fields in header or line items, they do not appear in the response JSON file.

Response Example

200 "Success" with barcode header field extraction

```
{
    "status":"DONE",
    "id":"2853a32c-9cf9-415f-9585-82c63c2fa699",
    "fileName":"qr_three_codes.pdf",
    "documentType":"invoice",
    "created":"2023-01-27T09:57:26.160906+00:00",
    "finished":"2023-01-27T09:58:20.383827+00:00",
    "clientId":"c_00",
    "languageCodes":[
        "de"
    ],
    "pageCount":1,
    "width":2480,
    "height":3507,
    "country":"MX",
    "bocrVersion":"1.7.0",
    "doxVersion":"local",
    "fileType":"pdf",
    "enrichment":{
        "sender":[

        ],
        "employee":[

        ],
        "product":[

        ]
    },
    "dataForRetrainingStatus":"notUsedForTraining",
    "extraction":{
        "headerFields":[
            {
                "name":"barcode",
                "category":"details",
                "value":"https://verificacfdi.facturaelectronica.sat.gob.mx/
default.aspx?
id=706220d0-3b0b-4801-82b8-5f771f8af9c1&re=CSA080218TQ8&rr=NME140730ME0&tt=000009
9576.720000&fe=ZI/I4A==",
                "rawValue":"https://verificacfdi.facturaelectronica.sat.gob.mx/
default.aspx?
id=706220d0-3b0b-4801-82b8-5f771f8af9c1&re=CSA080218TQ8&rr=NME140730ME0&tt=000009
9576.720000&fe=ZI/I4A==",
                "type":"string",
                "page":1,
```

```
            "confidence":1.0,
            "coordinates":{
                "x":0.14717741935483872,
                "y":0.262617621899059,
                "w":0.07782258064516129,
                "h":0.05503279155973767
            },
            "model":"ai",
            "group":1,
            "attributes":{
                "symbology":"QR"
            },
            "label":"barcode"
        },
        {
            "name":"barcode",
            "category":"details",
            "value":"https://verificacfdi.facturaelectronica.sat.gob.mx/
default.aspx?
id=706220d0-3b0b-4801-82b8-5f771f8af9c1&re=CSA080218TQ8&rr=NME140730ME0&tt=000009
9576.720000&fe=ZI/I4A==",
            "rawValue":"https://verificacfdi.facturaelectronica.sat.gob.mx/
default.aspx?
id=706220d0-3b0b-4801-82b8-5f771f8af9c1&re=CSA080218TQ8&rr=NME140730ME0&tt=000009
9576.720000&fe=ZI/I4A==",
            "type":"string",
            "page":1,
            "confidence":1.0,
            "coordinates":{
                "x":0.3294354838709677,
                "y":0.6854861705161106,
                "w":0.21129032258064517,
                "h":0.1497005988023952
            },
            "model":"ai",
            "group":2,
            "attributes":{
                "symbology":"QR"
            },
            "label":"barcode"
        },
        {
            "name":"barcode",
            "category":"details",
            "value":"https://verificacfdi.facturaelectronica.sat.gob.mx/
default.aspx?
id=706220d0-3b0b-4801-82b8-5f771f8af9c1&re=CSA080218TQ8&rr=NME140730ME0&tt=000009
9576.720000&fe=ZI/I4A==",
            "rawValue":"https://verificacfdi.facturaelectronica.sat.gob.mx/
default.aspx?
id=706220d0-3b0b-4801-82b8-5f771f8af9c1&re=CSA080218TQ8&rr=NME140730ME0&tt=000009
9576.720000&fe=ZI/I4A==",
            "type":"string",
            "page":1,
            "confidence":1.0,
            "coordinates":{
                "x":0.7411290322580645,
                "y":0.47619047619047616,
                "w":0.20725806451612902,
                "h":0.14656401482748788
            },
            "model":"ai",
            "group":3,
            "attributes":{
                "symbology":"QR"
            },
            "label":"barcode"
        },
```

```json
        {
            "name":"currencyCode",
            "category":"amounts",
            "value":"CHF",
            "rawValue":"",
            "type":"string",
            "page":1,
            "confidence":0.992719292640686,
            "coordinates":{
                "x":0.0,
                "y":0.0,
                "w":0.0,
                "h":0.0
            },
            "model":"ai",
            "label":"currencyCode"
        },
        {
            "name":"documentDate",
            "category":"document",
            "value":"2019-11-18",
            "rawValue":"19-11-18",
            "type":"date",
            "page":1,
            "confidence":0.9978566765785217,
            "coordinates":{
                "x":0.294758064516129,
                "y":0.16737952666096378,
                "w":0.08548387096774196,
                "h":0.00883946392928428
            },
            "model":"ai",
            "label":"documentDate"
        },
        {
            "name":"documentNumber",
            "category":"document",
            "value":"10101010",
            "rawValue":"10101010",
            "type":"string",
            "page":1,
            "confidence":0.947092592716217,
            "coordinates":{
                "x":0.3100806451612903,
                "y":0.18106643855146848,
                "w":0.07379032258064516,
                "h":0.00855431993156544
            },
            "model":"ai",
            "label":"documentNumber"
        },
        {
            "name":"grossAmount",
            "category":"amounts",
            "value":99576.72,
            "rawValue":"0000099576.720000",
            "type":"number",
            "page":1,
            "confidence":1.0,
            "coordinates":{
                "x":0.14717741935483872,
                "y":0.262617621899059,
                "w":0.07782258064516129,
                "h":0.05503279155973767
            },
            "model":"ai",
            "label":"grossAmount"
        },
```

```
            {
                "name":"receiverAddress",
                "category":"receiver",
                "value":"Max M\u00fcller, Teststral\u00dfe 99, 19012 Testort Rene
Teststral\u00dfe 100 19012 Testort",
                "rawValue":"Max M\u00fcller, Teststral\u00dfe 99, 19012 Testort Rene
Teststral\u00dfe 100 19012 Testort",
                "type":"string",
                "page":1,
                "confidence":0.7233287231620119,
                "coordinates":{
                    "x":0.567741935483871,
                    "y":0.20786997433704021,
                    "w":0.2270161290322581,
                    "h":0.053607071571143444
                },
                "model":"ai",
                "label":"receiverAddress"
            },
            {
                "name":"receiverContact",
                "category":"receiver",
                "value":"Rene M\u00fcller",
                "rawValue":"Rene M\u00fcller",
                "type":"string",
                "page":1,
                "confidence":0.5884397625923157,
                "coordinates":{
                    "x":0.5689516129032258,
                    "y":0.22783005417735958,
                    "w":0.08185483870967736,
                    "h":0.00855431993156544
                },
                "model":"ai",
                "label":"receiverContact"
            },
            {
                "name":"receiverName",
                "category":"receiver",
                "value":"Max M\u00fcller,",
                "rawValue":"Max M\u00fcller,",
                "type":"string",
                "page":1,
                "confidence":0.6223656535148621,
                "coordinates":{
                    "x":0.5677419304847717,
                    "y":0.2078699767589569,
                    "w":0.06330645084381104,
                    "h":0.008269175887107849
                },
                "model":"ai",
                "label":"receiverName"
            },
            {
                "name":"senderAddress",
                "category":"sender",
                "value":"Teststra\u00df\u00dfe 99 19012 Testort",
                "rawValue":"Teststra\u00df\u00dfe 99 19012 Testort",
                "type":"string",
                "page":1,
                "confidence":0.9818366663199499,
                "coordinates":{
                    "x":0.15201612903225806,
                    "y":0.08639863130881095,
                    "w":0.10443548387096777,
                    "h":0.02281151981750784
                },
                "model":"ai",
```

```
                "label":"senderAddress"
            },
            {
                "name":"senderBankAccount",
                "category":"sender",
                "value":"CHS1 5112 3453 5444 44",
                "rawValue":"CHS1 5112 3453 5444 44",
                "type":"string",
                "page":1,
                "confidence":0.6283774228323075,
                "coordinates":{
                    "x":0.06411290322580646,
                    "y":0.682919874536641,
                    "w":0.1411290322580645,
                    "h":0.007128599942971103
                },
                "model":"ai",
                "group":1,
                "label":"senderBankAccount"
            },
            {
                "name":"senderName",
                "category":"sender",
                "value":"Max M\u00fcller",
                "rawValue":"Max M\u00fcller",
                "type":"string",
                "page":1,
                "confidence":0.8665437601845373,
                "coordinates":{
                    "x":0.15201612903225806,
                    "y":0.0718562874251497,
                    "w":0.08225806451612905,
                    "h":0.009409751924721987
                },
                "model":"ai",
                "label":"senderName"
            },
            {

                "name":"taxId",
                "category":"amounts",
                "value":"CSA080218TQ8",
                "rawValue":"CSA080218TQ8",
                "type":"string",
                "page":1,
                "confidence":1.0,
                "coordinates":{
                    "x":0.14717741935483872,
                    "y":0.262617621899059,
                    "w":0.07782258064516129,
                    "h":0.05503279155973767
                },
                "model":"ai",
                "group":1,
                "label":"taxId"
            }
        ],
        "lineItems":[

        ]
    }
}
```

Extracted Header Fields and Extracted Line Items Categories

Fields can belong to a category. This is indicated by the category property of a field in the response JSON.
An example is a tax with multiple fields. Taxes are returned in the form of a category with the fields taxName,

taxRate, and taxAmount. See all field categories in Extracted Header Fields [page 291] and Extracted Line Items [page 300].

Response Example

400 "Bad Request"

```
{
   "code": "E93",
   "message": "Required parameters not provided.",
   "details": "string"
}
```

Response Example

401 "Unauthorized"

```
{
   "message": "No Authorization given in the request header"
}
```

Response Example

500 "Internal server error"

```
{
    "message": "Internal server error"
}
```

12.1.6.5 Save Ground Truth

Save the ground truth (correct values for document fields) for the specified document job ID.

This endpoint takes the job ID of a document submitted previously and returns the corresponding processing results, or an error, if the given ID isn't found.

Add to the payload `extraction` (list of all the extracted header fields and line items), and `enrichment` (list of the matched enrichment data).

For the fields, the following attributes are part of the ground truth:

- `name` (required)
- `value` (required)
- `rawValue` (optional)
- `page` (optional)
- `coordinates` (optional)

For enrichment data, the following attribute is part of the ground truth: `id` (required).

> ⚠ Caution
>
> It's technically possible to add other attributes to the ground truth payload (for example, `confidence`), but they have no impact on the stored values and are ignored.

> ⓘ **Note**
>
> After saving the ground truth of a document, the prediction confidence score of all header fields and line items is automatically set to 1.0 (100%). The service assumes that all field values are correct or have been manually corrected. Only save the ground truth of documents that have been reviewed and don't contain incorrect extraction results.

> ⚠ **Caution**
>
> It isn't possible to save ground truth if you used the `SAP_OCROnly_schema` for the document extraction. See second "Bad Request" error message in the *Response* section below.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/document/jobs/<id>`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`id`	Yes	String	*path*	The ID returned by the Upload Document [page 135] endpoint. For example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff
`payload`	Yes	JSON Object	*body*	Fields of the document (header fields and line items) and enrichment data
				ⓘ **Note** The structure of the `payload` is as the response returned by the Get Result [page 146] endpoint. However, while the top-N enrichment matches are returned in Get Result [page 146], for the *Save Ground Truth* endpoint, the enrichment list **must not** contain more than one (ground truth) match for each `sender` and `employee`.

Request Example

```
{
    "extraction":{
        "headerFields":[
            {
                "name":"documentDate",
```

```
                        "value":"2019-02-18"
                    },
                    {
                        "name":"grossAmount",
                        "value":200
                    }
                ],
                "lineItems":[
                    [
                        {
                            "name":"description",
                            "value":"Professional Services"
                        },
                        {
                            "name":"netAmount",
                            "value":200
                        },
                        {
                            "name":"unitPrice",
                            "value":200
                        },
                        {
                            "name":"materialNumber",
                            "value":"007"
                        }
                    ]
                ]
            },
            "enrichment":{
                "sender":[
                    {
                        "id":"BE0001"
                    }
                ],
                "employee":[
                    {
                        "id":"E0001"
                    }
                ]
            }
        }
}
```

Response

Response Fields

JSON Field	Description
message	Status message with information about the request
status	Status of the ground truth upload. Possible values: "PENDING", "DONE", or "FAILED"

The response is given as a status (201, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

201 "Created"

```
{
    "status": "DONE",
```

```
    "message": "Ground truth / corrected values uploaded successfully"
}
```

Response Example

400 "Bad Request"

```
{
  "code": "E93",
  "message": "Required parameters not provided.",
  "details": "string"
}
```

Response Example

400 "Bad Request" (with `SAP_OCROnly_schema`)

```
{
    "error":{
        "code":"ES068",
        "message":"Posting ground truth is not allowed for SAP_OCROnly_schema.",
        "details":[

        ]
    }
}
```

Response Example

401 "Unauthorized"

```
{
    "message": "No Authorization given in the request header"
}
```

Response Example

500 "Internal server error"

```
{
    "message":"Internal server error"
}
```

12.1.6.6 Confirm Document

Change the status of a document from "DONE" to "CONFIRMED". After that, the document status is permanent and cannot be changed anymore. The document extraction values cannot be changed anymore either. Also use this endpoint to enable the data feedback collection feature to allow documents to be used for retraining.

> ⓘ Note
>
> SAP reserves the right to use confirmed documents in the reporting of accuracy values and for analytics.
>
> If you set the parameter `dataForRetraining` to *true*, you allow the use of confirmed documents to retrain the machine learning models and improve the service.

Submitting retraining data and documents to SAP does not guarantee that SAP will actually use the data for improving the service, or that SAP guarantees that potential errors will be fixed in future improved versions of the service.

The prediction confidence score of all header fields and line items is automatically set to 1.0 (100%) for confirmed documents. The service assumes that all field values are correct or have been manually corrected. Only confirm documents that have been reviewed and don't contain incorrect extraction results.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/document/jobs/<id>/confirm`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`dataForRetraining`	No	Boolean	*query*	Set to *true* to allow confirmed documents to be used to retrain the service's machine learning models. Set to *false* if you do not want to use the data feedback collection feature.
				ⓘ Note The data feedback collection feature is only available on production environments for enterprise accounts. This feature is not available for trial account users. SAP reserves the right to reject documents submitted by the customers for retraining.
				To use the data feedback collection feature, Create Configuration [page 122] setting the `dataFeedbackCollection` configuration key to *true*.
`id`	Yes	String	*path*	The ID returned by the Upload Document [page 135] endpoint. For example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff.

Response

Response Fields

JSON Field	Description
message	Status message with information about the request
status	Document confirmation status. Possible value: "CONFIRMED"

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "status": "CONFIRMED",
    "message": "Document confirmed successfully."
}
```

400 " Bad Request" (dataFeedbackCollection configuration key is not set to *true*)

```
{
    "message": "Data feedback collection is only possible with the correct tenant
configuration. Please set dataFeedbackCollection to true."
}
```

12.1.6.7 Get Document File

Get the original document file you uploaded to the service.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: /document-information-extraction/v1

URL Endpoint Path: /document/jobs/<id>/file

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
id	Yes	String	*path*	The ID returned by the Upload Document [page 135] endpoint. For example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff.

Response

The response is given as a status (200, 400, 401, 404, or 500) and document file in the format previously uploaded using the Upload Document [page 135] endpoint. See Common Status and Error Codes [page 235].

12.1.6.8 Get All Pages Text

Get the text of all pages of a document.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/document/jobs/<id>/pages/text`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
id	Yes	String	*path*	The ID returned by the Upload Document [page 135] endpoint. For example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff

Response

Response Fields

JSON Field	Description
results	List containing the text and the corresponding bounding boxes (specified by the returned coordinates) of all pages of a document

The response is given as a status (200, 400, 401, 404, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "results":{
        "1":[
            {
```

```
            "word_boxes":[
                {
                    "bbox":[
                        [
                            890,
                            141
                        ],
                        [
                            1028,
                            174
                        ]
                    ],
                    "content":"Rocket"
                },
                {
                    "bbox":[
                        [
                            1049,
                            141
                        ],
                        [
                            1275,
                            182
                        ]
                    ],
                    "content":"Enterprises"
                },
                {
                    "bbox":[
                        [
                            1297,
                            143
                        ],
                        [
                            1365,
                            183
                        ]
                    ],
                    "content":"Pty"
                },
                {
                    "bbox":[
                        [
                            1383,
                            140
                        ],
                        [
                            1443,
                            174
                        ]
                    ],
                    "content":"Ltd"
                }
            ],
            "bbox":[
                [
                    890,
                    140
                ],
                [
                    1443,
                    184
                ]
            ]
        }
    ]
    }
}
```

12.1.6.9 Get Single Page Text

Get the text of a single page of a document.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/document/jobs/<id>/pages/<no>/text`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`id`	Yes	String	*path*	The ID returned by the Upload Document [page 135] endpoint. For example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff
`no`	Yes	String	*path*	The page number of the document

Response

Response Fields

JSON Field	Description
`value`	List containing the text and the corresponding bounding boxes (specified by the returned coordinates) of a single page of a document

The response is given as a status (200, 400, 401, 404, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "value":[
        {
            "word_boxes":[
                {
                    "bbox":[
                        [
                            890,
                            141
                        ],
                        [
                            1028,
                            174
```

```
                              ]
                          ],
                          "content":"Rocket"
                      },
                      {
                          "bbox":[
                              [
                                  1049,
                                  141
                              ],
                              [
                                  1275,
                                  182
                              ]
                          ],
                          "content":"Enterprises"
                      },
                      {
                          "bbox":[
                              [
                                  1297,
                                  143
                              ],
                              [
                                  1365,
                                  183
                              ]
                          ],
                          "content":"Pty"
                      },
                      {
                          "bbox":[
                              [
                                  1383,
                                  140
                              ],
                              [
                                  1443,
                                  174
                              ]
                          ],
                          "content":"Ltd"
                      }
                  ],
                  "bbox":[
                      [
                          890,
                          140
                      ],
                      [
                          1443,
                          184
                      ]
                  ]
              }
          ]
      }
```

12.1.6.10 Get Request Options

Get the request options for a document.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/document/jobs/<id>/request`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`id`	Yes	String	*path*	The ID returned by the Upload Document [page 135] endpoint. For example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff

Response

Response Fields

JSON Field	Description
`documentType`	Type of the document submitted
`extraction`	Dictionary containing all the extracted header fields and line items
`receivedDate`	The date when the document was received, for example, 2020-02-17.

The response is given as a status (200, 400, 401, 404, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
  "extraction": "...",
  "documentType": "invoice",
  "receivedDate": "2020-02-17"
}
```

12.1.6.11 Get Templates Associated with Document

Get all the templates associated with the specified document ID.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/document/jobs/<id>/template`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`clientId`	Yes	String	*query*	The ID of the client. For example: c_00
`id`	Yes	String	*path*	The ID returned by the Upload Document [page 135] endpoint. For example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff

Response

Response Fields

JSON Field	Description
`templateId`	The ID of the template associated with the document ID.

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "templateId":[
        "4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff"
    ]
}
```

12.1.6.12 Delete Document

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/document/jobs`

HTTP Method: *DELETE*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
payload	Yes	JSON Object	*body*	List of document IDs you want to delete

Response

Response Fields

JSON Field	Description
message	Status message with information about the request
processedTime	Timestamp in RFC format
status	Deletion status of the document. Possible values: "PENDING", "DONE", or "FAILED"

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
  "status": "DONE",
  "message": "Documents deleted successfully.",
  "processedTime": "2020-03-26T17:00:00.000000+00:00"
}
```

12.1.7 Enrichment Data API

Document Information Extraction can also enrich the information extracted from documents with your existing structured data (typically master data records).

Enrichment in that context means to provide additional information to a document, which is not directly contained on a document, but which is inferred based on information, which is contained on a document in conjunction to other external data.

You can, for example, infer the proprietary ID of a customer from another SAP system based on the sender address contained on an invoice document. Even though the customer ID is not explicitly contained on the invoice, the ID from the SAP system can be inferred by using the address data contained on the invoice by matching it against the relevant master data.

The service matches enrichment data entities with the Extracted Header Fields [page 291] and Extracted Line Items [page 300] from processed documents.

The Enrichment Data API provides the functionalities to create, update, get and delete enrichment data. After enrichment data entities have been maintained, please check the usage of the enrichment property in Upload Document [page 135] in order to leverage the matching of enrichment data to extracted fields.

The Enrichment Data API consists of the following endpoints:

- Create Enrichment Data [page 175]
- List Data-Persistence Jobs [page 182]
- Get Enrichment Data [page 184]
- Get Enrichment Data Creation or Deletion Status [page 186]
- Create Data Activation [page 187]
- Get Data Activation Details [page 188]
- Delete Enrichment Data (Synchronous) - Deprecated [page 189]
- Delete Enrichment Data (Asynchronous) [page 191]

Related Information

Data Enrichment: Best Practices [page 284]
Enrichment Parameter [page 141]

12.1.7.1 Create Enrichment Data

Create or update one or more enrichment data entities.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/data/jobs`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	query	The ID of the client. For example: c_00.
payload	Yes	JSON Object	body	List containing enrichment data entities in value property. The entities can be: • BusinessEntity [page 178] • Employee [page 180] • Product [page 180] See request examples below. See also Data Variants [page 181].
type	Yes	String	query	The type of enrichment data entities used for matching specified in the JSON string of the payload. Available values: businessEntity, employee, and product.
subtype	No	String	query	The subtype of enrichment data entities used for matching with type **businessEntity** specified in the JSON string of the payload. Available values: supplier, customer, and companyCode.

Request Examples

Create BusinessEntity [page 178] entities:

```
payload:
{
    "value":[
        {
            "id":"BE0001",
            "name":"Emma Dowerg",
            "accountNumber":"SK2421",
            "address1":"Amalie-Klemm-Platz 0/9, 48581, Geithain",
            "address2":"Near city church",
            "city":"Geithain",
            "countryCode":"DE",
            "postalCode":"48581",
            "state":"Schleswig-Holstein",
            "email":"e.dowerf@mustermail.com",
            "phone":"+49(0) 909979463",
            "bankAccount":"DE345982837402",
            "taxId":"DE435531312"
        },
        {
            "id":"BE0002",
            "name":"Ioannis Kruschwitz",
            "accountNumber":"393H292",
            "address1":"Alina-Reichmann-Allee 73, 63228, Staßfurt",
            "city":"Staßfurt",
            "countryCode":"DE",
            "postalCode":"63228",
            "state":"Hessen",
            "email":"Ioannis.Kruschwitz@mustermail.com",
            "phone":"+49(0) 818172710",
            "bankAccount":"DE1093628093743",
            "taxId":"DE593029048"
```

```
        }
    ]
}
type: businessEntity
clientId: c_00
subtype: supplier
```

Create Employee [page 180] entities:

```
payload:
{
    "value":[
        {
            "id":"E0001",
            "email":"john.will.doe@mustermail.com",
            "firstName":"John",
            "middleName":"William",
            "lastName":"Doe"
        },
        {
            "id":"E0002",
            "email":"m.gierschner@mustermail.com",
            "firstName":"Maren",
            "middleName":"Volkhard",
            "lastName":"Gierschner"
        }
    ]
}
type: employee
clientId: c_00
```

Create Product [page 180] entities:

```
payload:
{
  "value": [
      {
        "id": "12342",
        "description": "Glycerin Retinol 80 ML",
        "materialNumber": "B676817",
        "unitPrice": "1000,0 €",
        "unitOfMeasure": "LTR"
      }
    ]
}
type: product
clientId: c_00
```

Response

Response Fields

JSON Field	Description
id	Request ID
status	Status of the request. Possible values: "PENDING", "DONE", or "FAILED"

The response is given as a status (201, 400, 401, 422, 429, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

201 "Success"

```
{
  "id": "484b6e1c-501c-4a07-85cb-84554656a175",
  "status": "PENDING"
}
```

Related Information

12.1.7.1.1 Entities

Entities are several actors which can be addressed by a business document. A business entity can be, for example, a customer and a supplier. The employee entity represents an employee in the company. The product entity represents a specific good or service available in a catalog or system.

Related Information

12.1.7.1.1.1 BusinessEntity

A `businessEntity` can represent different kind of organizations with which you deal as a company. It can represent, for example, suppliers and customers.

See to create `businessEntity` entities.

Key	Type	Length (maximum length of the string)	Description	Example
accountNumber	String	100	Account number of the business entity. This refers to a business account number and not a bank account number.	1213414
address1	String	150	Complete address fields of the business entity. Use a comma (",") to separate each individual field of the address.	Musterstraße 21, 13123, Musterstadt
address2	String	100	Any additional fields or landmarks that are part of the address.	Near Stadt Dom
bankAccount	String	100	Bank account number of the business entity. Enter the bank account number in a continuous string without spaces.	DE32245443233323
city	String	100	City of the business entity.	Musterstadt
countryCode	String	100	Country/Region of the business entity.	Deutschland
id	String	100	Unique identifier of the business entity in the user system.	BE21e112
email	String	100	Email address of the business entity.	mustermann@muster-mail.com
name	String	256	Name of the business entity.	Muster Mann GmbH
phone	String	100	Phone number of the business entity. Add the country/region code with the "+" symbol before the phone number.	+49131231331
postalCode	String	100	Postal code of the business entity.	12323
state	String	100	State of the business entity.	Rhineland Palatinate
taxId	String	100	Tax ID of the business entity.	DE123456789

12.1.7.1.1.2 Employee

An `employee` is a specific employee within the company.

See Create Enrichment Data [page 175] to create `employee` entities.

Key	Type	Length (maximum length of the string)	Description	Example
email	String	256	Email address of the employee.	m.gierschner@muster-mail.com
firstName	String	100	First name of the employee.	Maren
id	String	128	Unique identifier of the employee in the user system.	E0002
lastName	String	100	Last name of the employee.	Gierschner
middleName	String	100	Middle name of the employee.	Volkhard

12.1.7.1.1.3 Product

A `product` is a specific good or service available in a catalog or system.

See Create Enrichment Data [page 175] to create `product` entities.

Key	Type	Length (maximum length of the string)	Description	Example
description	String	100	Description of the product.	Glycerin Retinol 80 ML
id	String	128	Unique identifier of the product in the user system.	12342
materialNumber	String	100	Unique code that identifies a specific good or service in a supplier catalog or system.	B676817
unitOfMeasure	String	100	The unit of measure UN/CEFACT code.	LTR for liter and KGM for kilogram.
unitPrice	String	100	Price for a single instance of an object.	1000,0 €

12.1.7.1.2 Data Variants

Use variants to create multiple versions of the same data record, which all point to the same record ID.

To create a data record variant, add the `variant` key to the Create Enrichment Data [page 175] payload:

```
payload:
{
    "value":[
        {
            "id":"BE0001",
            "name":"Emma Dowerg",
            "address1":"Amalie-Klemm-Platz 0/9, 48581, Geithain",
            "bankAccount":"DE345982837402",
            "taxId":"DE435531312",
            "variant":"2"
        }
    ]
}
type: businessEntity
clientId: c_00
subtype: supplier
```

All the variants are used for the enrichment. If a data record match is associated with a variant ID, the matched variant ID is returned by Get Result [page 146] alongside the usual enrichment result information. For example:

```
enrichment: {
    "id":"BE0001",
    "confidence":98.647,
    "variant":2
}
```

The variant ID is an optional parameter. If absent, the data record is not associated to any variant. If used, variant IDs can be a number in the inclusive range 1 - 9. Any other variant ID is invalid and will result in an error.

Creating another master data record with the same ID and variant ID will not result in an error. Instead, the behavior is the same as creating a data record with an already existing ID, but both without variant IDs. See Data Duplicates [page 182].

> ⓘ Note
>
> A single invalid variant ID value (for example, a variant that is not a number in the inclusive range 1 - 9) will cause the whole batch (API request) to fail.

> → Tip
>
> You can create multiple variants of the same data record (all sharing the same ID) but in different languages.

12.1.7.1.3 Data Duplicates

Find out how how the Document Information Extraction service handles the upload of duplicated master data records.

What are data duplicates?

A master data record "X" is considered a duplicate by the Document Information Extraction service if there is another existing record "Y" which fulfills all of the following conditions:

- "X" has the exact same ID as "Y".
- "X" has the exact same variant ID as "Y". If both records have no variant ID they are also considered as equals.
- "X" and "Y" are created from the same tenant, client and service instance.

How does Document Information Extraction handle duplicates?

The service filters out duplicate records as part of the automatic or manual data activation. If one or more duplicates are identified, the following update rule is applied to all of them: the most recently created record replaces all previously created versions of that record.

This process optimizes the service experience and results for most common use cases in which duplicated records are not intended. If duplicated records are required as part of an individual use case, this can be achieved using variant IDs.

12.1.7.2 List Data-Persistence Jobs

Returns a list of all data-persistence jobs for this tenant.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/data/jobs`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	No	String	*query*	The ID of the client. For example: c_00
limit	No	Integer	*query*	Items per page. Defines a maximum limit. For example: 10. See Technical Constraints [page 288].
offset	No	Integer	*query*	Offset of the first item to be returned. For example: 10
order	No	String	*query*	Order criteria for the retrieved data-persistence jobs. Possible values: created, client, or status. For example: created asc (sorts by creation date in ascending order)
status	No	String	*query*	The status of this data-persistence job. Possible values: "PENDING", "SUCCESS", or "FAILED"

Response

Response Fields

JSON Field	Description
clientId	ID of the client this data-persistence job was created for
created	Time when the data-persistence job was created
id	Data-persistence job ID
status	The status of this data-persistence job. Possible values: "PENDING", "SUCCESS", or "FAILED"

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "value":{
        "id":"c4f25368-d3e6-43f7-a0b4-55adf7f54e95",
        "status":"PENDING",
        "clientId":"c_00",
        "created":"2020-05-08T10:39:59.916359+00:00"
    }
}
```

12.1.7.3 Get Enrichment Data

Retrieve one or more enrichment data entities.

> ⓘ Note
>
> Enrichment data is refreshed automatically every 4 hours. It might take up to 4 hours until the enrichment data prediction is available in the Get Result [page 146] response. Manual data activation is also available and is the recommended process. You can set data activation to manual using the following endpoints:
>
> - Create Configuration [page 122]
> - Create Data Activation [page 187]

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/data`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. For example: c_00
companyCode	No	String	*query*	The company code of a single entry
id	No	String	*query*	The data ID of a single entry
limit	No	Integer	*query*	Items per page. Defines a maximum limit. For example: 10. See Technical Constraints [page 288].
offset	No	Integer	*query*	Offset of the first item to be returned. For example: 10
type	Yes	String	*query*	The type of enrichment data entities used for matching. Available values: `businessEntity`, `employee`, and `product`.
subtype	No	String	*query*	The subtype of enrichment data entities used for matching with type **businessEntity**. Available values: `supplier`, `customer`, and `companyCode`.
system	No	String	*query*	The system of a single entry

Response

Response Fields

JSON Field	Description
accountNumber	Account number of the enrichment data entity
address1	Address of the enrichment data entity
address2	Additional address of the enrichment data entity
bankAccount	Bank account number of the enrichment data entity
city	City name of the enrichment data entity
companyCode	Company code of the enrichment data entity
countryCode	Country/Region code of the enrichment data entity
email	Email address of the enrichment data entity
id	ID of the enrichment data entity
name	Name of the enrichment data entity
phone	Phone number of the enrichment data entity
postalCode	Postal code of the enrichment data entity
state	State code of the enrichment data entity
system	System of the enrichment data entity
taxId	Tax ID of the enrichment data entity

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
  "value": [
    {
      "id": "BE0001",
      "name": "A",
      "accountNumber": "12345",
      "address1": "A street 5",
      "address2": "",
      "city": "Heidelberg",
      "countryCode": "DE",
      "postalCode": "69117",
      "state": "BW",
      "email": "a@a.com",
      "phone": "",
      "bankAccount": "000001",
      "taxId": "999",
      "companyCode": "4711",
      "system": "System A"
    }
  ]
}
```

12.1.7.4 Get Enrichment Data Creation or Deletion Status

Give a data persistence job ID to check the database and receive information on this data persistence job.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/data/jobs/`<id>

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
id	Yes	String	*path*	The ID returned by the Create Enrichment Data [page 175] or Delete Enrichment Data (Asynchronous) [page 191] endpoints. For example: 29812f26-464e-4ee6-be63-731859cf99f3.

Response

Response Fields

JSON Field	Description
id	Request ID.
processedTime	Amount of time it took to process the request.
refreshedAt	Date in extended ISO 8601 format (for example, "2021-01-16T13:36:29.453713+00:00"). It tells when the enrichment data job was refreshed for the last time. When the response is "**null**", it means that the enrichment data has not yet been refreshed.
status	Status of the request. Possible values: "PENDING", "DONE", or "FAILED".

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "value":{
        "id":"b89645b4-605b-45cd-bf69-1147875e75f5",
        "status":"SUCCESS",
        "processedTime":"0:00:00.063022",
```

```
        "refreshedAt":"2021-01-16T13:36:29.453713+00:00"
    }
}
```

Response Example

400 "Bad Request"

```
{
  "code": "E5",
  "message": "Failed to retrieve data.",
  "details": "string"
}
```

Response Example

401 "Unauthorized"

```
{
  "message": "No Authorization given in the request header"
}
```

12.1.7.5 Create Data Activation

Create a data activation job record to see new or updated enrichment data in the extraction results if you are using the manual data activation process. Only activated enrichment data will be added to the extraction results.

> → Remember
>
> Before creating an enrichment data activation job record, you need to Create Configuration [page 122].

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/data/activation`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
type	No	String	*query*	The type of enrichment data entities used for matching. Available values: `businessEntity`, `employee`, and `product`.

Parameter	Required	Data Type	Parameter Type	Description
subtype	No	String	*query*	The subtype of enrichment data entities used for matching with type **businessEntity**. Available values: supplier, customer, and companyCode.

Response

Response Fields

JSON Field	Description
id	ID of the enrichment data activation job record
status	Status of the request. Possible values: "PENDING", "DONE", or "FAILED"

The response is given as a status (201, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

201 "Success"

```
{
  "id": "484b6e1c-501c-4a07-85cb-84554656a175",
  "status": "PENDING"
}
```

12.1.7.6 Get Data Activation Details

Give an enrichment data activation job record ID to check the database, and receive information on this data activation job.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: /document-information-extraction/v1

URL Endpoint Path: /data/activation/<id>

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
id	Yes	String	*path*	The ID returned by the Create Data Activation [page 187] endpoint. For example: 484b6e1c-501c-4a07-85cb-84554656a175.

Response

Response Fields

JSON Field	Description
created	Time when the enrichment data was submitted for processing
finished	Time when the enrichment data status changed to "DONE", or "FAILED"
id	ID of the enrichment data activation job record
processedTime	Timestamp in RFC format
status	Status of the request. Possible values: "PENDING", "DONE", or "FAILED"

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
  "value": {
    "id": "484b6e1c-501c-4a07-85cb-84554656a175",
    "status": "DONE",
    "processedTime": "0:01:00",
    "created": "2019-07-04T15:20:37.668873+00:00",
    "finished": "2019-07-04T15:21:37.668873+00:00"
  }
}
```

12.1.7.7 Delete Enrichment Data (Synchronous) - Deprecated

Perform synchronous deletion of existing data records for specified fields.

> ⚠ Caution
>
> This endpoint has been deprecated and is scheduled for decommissioning in November 2024. Please use the endpoint Delete Enrichment Data (Asynchronous) [page 191] to delete data records.

> ⓘ Note
>
> To delete large numbers of data records, use **only** the endpoint Delete Enrichment Data (Asynchronous) [page 191].

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/data`

HTTP Method: *DELETE*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`clientId`	Yes	String	*query*	The ID of the client. For example: c_00
`payload`	Yes	JSON Object	*body*	Comma-separated list of data record IDs that you want to delete
`type`	Yes	String	*query*	The type of enrichment data entities used for matching specified in the JSON string of the `payload`. Available values: `businessEntity`, `employee`, and `product`.
`subtype`	No	String	*query*	The subtype of enrichment data entities used for matching with type **businessEntity** specified in the JSON string of the `payload`. Available values: `supplier`, `customer`, and `companyCode`.

Response

Response Fields

JSON Field	Description
`deleted`	Total number of data records deleted with this request

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
```

```
    "deleted": "2"
}
```

12.1.7.8 Delete Enrichment Data (Asynchronous)

Perform asynchronous deletion of existing data records for specified fields.

This endpoint accepts an array of data record IDs that you want to delete. If no array is entered in the `payload`, all entries are deleted.

You can also delete large numbers of data records for all clients per data type (businessEntity, employee, or product) by entering only the `type` parameter in your request. If you do not specify `clientId` and `type`, you will delete all data records for a tenant.

> → Tip
>
> Delete outdated and no longer used data records frequently to improve the performance of the data enrichment feature when matching a business document to an enrichment data record based on the information extracted from the document.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/data/jobs`

HTTP Method: *DELETE*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	No	String	*query*	The ID of the client. For example: c_00
payload	Yes	JSON Object	*body*	Comma-separated list of data record IDs that you want to delete. All data records are deleted if `payload` is empty.
type	No	String	*query*	The type of enrichment data entities used for matching specified in the JSON string of the `payload`. Available values: `businessEntity`, `employee`, and `product`.

Parameter	Required	Data Type	Parameter Type	Description
subtype	No	String	*query*	The subtype of enrichment data entities used for matching with type **businessEntity** specified in the JSON string of the `payload`. Available values: `supplier`, `customer`, and `companyCode`.

Request Examples

Delete all data records:

```
payload:
{
    "value":[]
}
```

Delete all BusinessEntity [page 178] data records:

```
payload:
{
    "value":[]
}
type: businessEntity
```

Delete all Employee [page 180] data records:

```
payload:
{
    "value":[]
}
type: employee
```

Delete all Product [page 180] data records:

```
payload:
{
    "value":[]
}
type: product
```

Response

Response Fields

JSON Field	Description
id	Request ID
status	Status of the request. Possible values: "PENDING", "DONE", or "FAILED"

The response is given as a status (201, 400, 401, 422, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

201 "Success"

```
{
    "id": "484b6e1c-501c-4a07-85cb-84554656a175",
    "status": "PENDING"
}
```

12.1.8 Schema API

Create schemas containing data fields found in standard or custom document types. You can use these schemas as a basis for creating templates. You can select schemas and associated templates when adding documents. The Schema API provides endpoints to create, list, update, and delete schemas and schema versions.

The Schema API consists of the following endpoints:

- Create Schema [page 193]
- Get Schema [page 195]
- Get Schema Capabilities [page 197]
- Update Schema [page 199]
- Get Schema Details [page 200]
- Delete Schema [page 204]
- Create Schema Version [page 205]
- Activate Schema Version [page 206]
- Deactivate Schema Version [page 207]
- Add Fields to Schema Version [page 208]
- Get Schema Versions [page 215]
- Get Schema Version Details [page 216]
- Delete Schema Versions [page 218]

12.1.8.1 Create Schema

Create one or more schemas for a client.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/schemas`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
payload	Yes	JSON Object	*body*	List containing `clientId`, `name`, `schemaDescription`, `documentType`, and `documentTypeDescription`.

> → Remember
>
> Each `schemaDescription` can have a maximum length of 500 characters.

> → Tip
>
> To get better results for extractions with generative AI (available with the service plan **Document Information Extraction, premium edition** only), you can use the `schemaDescription` value. This value is used in the prompt for document processing with large language models (LLMs).

Request Example

```
{
    "clientId":"c_00",
    "name":"Custom_Payment_Advice_Schema",
    "schemaDescription":"Schema For Accounts Department Payment Advices",
    "documentType":"paymentAdvice",
    "documentTypeDescription":"Payment Advice with Order Number"
}
```

Response

Response Fields

JSON Field	Description
created	Time when the schema was created.
id	ID of the schema

The response is given as a status (201, 400, 401, 429, or 500) and JSON file. See Common Status and Error Codes [page 235] and Technical Constraints [page 288].

Response Example

201 "Success"

```
{
    "id":"484b6e1c-501c-4a07-85cb-84554656a175",
    "created":"2020-03-26T17:00:00.000000+00:00"
}
```

12.1.8.2 Get Schema

Retrieve all schemas for a client.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/schemas`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client used when creating the schema. Example: c_00
documentType	No	String	*query*	The type of the document used when creating the schema. For example: `custominvoice`, or `paymentAdvice`
limit	No	Integer	*query*	Maximum number of schemas to be returned (maximum allowed value: 1000). For example: 10
offset	No	Integer	*query*	Index of the first schema to be retrieved. For example: 20
order	No	String	*query*	Order criteria of schemas to be returned. For example: "name asc", (sorts by name in ascending order)
predefined	No	Boolean	*query*	Set to true for standard documents or false for custom documents.

Response

Response Fields

JSON Field	Description
`created`	Time when the schema was created
`documentType`	Type of the document submitted
`documentTypeDescription`	Description of the document submitted
`id`	ID of the schema
`name`	Name of the schema
`predefined`	True for standard documents, false for custom documents
`schemaDescription`	Description of the schema
`state`	State of the schema
`updated`	Time when the schema was updated

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "schemas":[
        [
            {
                "name":"Basic Involve FormatSchema",
                "schemaDescription":"SAP Invoice Schema",
                "documentType":"Invoice",
                "documentTypeDescription":"Payment Advice with Order Number",
                "id":"484b6e1c-501c-4a07-85cb-84554656a175",
                "predefined":"True",
                "created":"2020-03-26T17:00:00.000000+00:00",
                "updated":"2020-04-26T17:00:00.000000+00:00",
                "state":"draft"
            },
            {
                "name":"Daimier Payment Advice Schema",
                "schemaDescription":"Payment Advice Schema",
                "documentType":"Payment Advice",
                "documentTypeDescription":"Payment Advice with Order Number",
                "id":"484b6e1c-501c-4a07-85cb-84554656a189",
                "predefined":"False",
                "created":"2020-03-26T17:00:00.000000+00:00",
                "updated":"2020-04-26T17:00:00.000000+00:00",
                "state":"active"
            }
        ]
    ]
}
```

12.1.8.3 Get Schema Capabilities

Retrieve all schema capabilities.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/schemas/capabilities`

HTTP Method: *GET*

Request Parameters

No parameters

Response

Response Fields

JSON Field	Description
`documentTypes`	List of the document types that are relevant to schemas
`formatting`	List of the possible formatting for schemas
`setupTypes`	List of the possible setup types for schemas
`state`	List of the possible statuses for schemas

The response is given as a status (200, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "documentTypes":[
        "invoice",
        "paymentAdvice",
        "purchaseOrder",
        "custom",
        "businessCard"
    ],
    "state":[
        "active",
        "inactive",
        "draft"
    ],
    "setupTypes":[
        {
```

```
                "name":"static",
                "properties":[

                ]
            },
            {
                "name":"ml",
                "properties":[
                    "x",
                    "y",
                    "w",
                    "z"
                ]
            },
            {
                "name":"...",
                "properties":"[]"
            }
        ],
        "formatting":[
            {
                "name":"string",
                "properties":[
                    {
                        "name":"length",
                        "values":[
                            "number"
                        ]
                    }
                ]
            },
            {
                "name":"number",
                "properties":[
                    {
                        "name":"length",
                        "values":[
                            "number"
                        ]
                    },
                    {
                        "name":"thousandSeparator",
                        "values":[
                            ".",
                            ",",
                            " "
                        ]
                    },
                    {
                        "name":"decimalSeparator",
                        "values":[
                            ".",
                            ",",
                            " "
                        ]
                    }
                ]
            },
            {
                "name":"...",
                "properties":"[]"
            }
        ]
}
```

12.1.8.4 Update Schema

Update existing schemas for a client.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/schemas/``<schemaId>`

HTTP Method: *PUT*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`clientId`	Yes	String	*query*	The ID of the client used when creating the schema. Example: c_00
`payload`	Yes	JSON Object	*body*	List containing name, `schemaDescription`, and `documentTypeDescription`.
`schemaId`	Yes	String	*path*	The ID returned by the endpoint Create Schema [page 193]. Example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff

Request Example

```
{
    "name":"Custom_Payment_Advice_Schema",
    "schemaDescription":"Schema For Accounts Department Payment Advices",
    "documentTypeDescription":"Payment Advice with Order Number"
}
```

Response

Response Fields

JSON Field	Description
`message`	Status message with information about the request

The response is given as a status (201, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

201 "Success"

```
{
    "message":"Schema has been updated successfully."
}
```

12.1.8.5 Get Schema Details

Retrieve schema details for a client.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/schemas/<schemaId>`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client used when creating the schema. Example: c_00
schemaId	Yes	String	*path*	The ID returned by the endpoint Create Schema [page 193]. Example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff

Response

Response Fields

JSON Field	Description
created	Time when the schema was created
documentType	Type of the document used for the schema
documentTypeDescription	Description of the document used for the schema
headerFields	List of header fields that are part of the schema
id	ID of the schema

JSON Field	Description
lineItemFields	List of line items that are part of the schema
name	Name of the schema
predefined	True for standard documents, false for custom documents
schemaDescription	Description of the schema
state	State of the schema
updated	Time when the schema was updated

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```json
{
    "name":"Basic Involve FormatSchema",
    "schemaDescription":"SAP Invoice Schema",
    "documentType":"Invoice",
    "documentTypeDescription":"Payment Advice with Order Number",
    "id":"484b6e1c-501c-4a07-85cb-84554656a175",
    "created":"2020-03-26T17:00:00.000000+00:00",
    "updated":"2020-04-26T17:00:00.000000+00:00",
    "predefined":"FALSE",
    "state":"draft",
    "headerFields":[
        {
            "name":"GrossAmountValue",
            "description":"TotalAmountValue",
            "defaultExtractor":{

            },
            "setupType":"static",
            "setupTypeVersion":"1.0.0",
            "setup":{
                "type":"default",
                "priority":1,
                "filter":[
                    {
                        "key":"language",
                        "value":"EN"
                    },
                    {
                        "key":"language",
                        "value":"DE"
                    }
                ],
                "properties":[
                    {
                        "key":"deploymentID",
                        "value":"123e4567-e89b-12d3-a456-426614174000."
                    },
                    {
                        "key":"fieldName",
                        "value":"GrossAmount"
                    }
                ]
            },
            "formattingType":"number",
            "formatting":{
```

```
            "length":"64",
            "precision":"3",
            "decimalSeparator":".",
            "thousandSeparator":","
        },
        "formattingTypeVersion":"1.0.0"
    },
    {
        "name":"sendersFullName",
        "description":"Name of Sender",
        "defaultExtractor":{
            "fieldName":"senderName"
        },
        "setup":{
            "type":"default",
            "priority":1,
            "filter":[
                {
                    "key":"language",
                    "value":"EN"
                },
                {
                    "key":"language",
                    "value":"DE"
                }
            ],
            "properties":[
                {
                    "key":"deploymentID",
                    "value":"123e4567-e89b-12d3-a456-426614174000."
                },
                {
                    "key":"fieldName",
                    "value":"senderName"
                }
            ]
        },
        "setupTypeVersion":"",
        "setupType":"",
        "formattingType":"",
        "formatting":{

        },
        "formattingTypeVersion":"1.0.0"
    }
],
"lineItemFields":[
    {
        "name":"Amount",
        "description":"TotalAmountValue",
        "defaultExtractor":{

        },
        "setupType":"static",
        "setupTypeVersion":"1.0.0",
        "setup":{
            "type":"default",
            "priority":1,
            "filter":[
                {
                    "key":"language",
                    "value":"EN"
                },
                {
                    "key":"language",
                    "value":"DE"
                }
            ],
```

```json
            "properties":[
                {
                    "key":"deploymentID",
                    "value":"123e4567-e89b-12d3-a456-426614174000."
                },
                {
                    "key":"fieldName",
                    "value":"NetAmount"
                }
            ]
        },
        "formattingType":"number",
        "formatting":{
            "length":"64",
            "precision":"3",
            "decimalSeparator":".",
            "thousandSeparator":","
        },
        "formattingTypeVersion":"1.0.0"
    },
    {
        "name":"WithdrawalDate",
        "description":"Date of Withdrawal",
        "defaultExtractor":{

        },
        "setupType":"static",
        "setupTypeVersion":"1.0.0",
        "setup":{
            "type":"default",
            "priority":1,
            "filter":[
                {
                    "key":"language",
                    "value":"EN"
                },
                {
                    "key":"language",
                    "value":"DE"
                }
            ],
            "properties":[
                {
                    "key":"deploymentID",
                    "value":"123e4567-e89b-12d3-a456-426614174000."
                },
                {
                    "key":"fieldName",
                    "value":"DocumentDate"
                }
            ]
        },
        "formattingType":"date",
        "formatting":{
            "dateformat":"dd/mm/yy"
        },
        "formattingTypeVersion":"1.0.0"
    }
    ]
}
```

12.1.8.6 Delete Schema

Delete one or more schemas for a client.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/schemas`

HTTP Method: *DELETE*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client used when creating the schema. Example: c_00
payload	Yes	JSON Object	*body*	Comma-separated list of `schemaIds` you want to delete.

Request Example

```
{
    "value":[
       "4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff"
    ]
}
```

Response

Response Fields

JSON Field	Description
message	Status message with information about the request

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "message":"Schemas deleted successfully."
}
```

12.1.8.7 Create Schema Version

Create a new version for a schema.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/schemas/`<code style="color:purple"><schemaId></code>

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`clientId`	Yes	String	*query*	The ID of the client used when creating the schema. Example: c_00
`schemaId`	Yes	String	*path*	The ID returned by the endpoint Create Schema [page 193]. Example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff

Response

Response Fields

JSON Field	Description
`created`	Time when the schema version was created.
`id`	ID of the schema
`version`	Version of the schema

The response is given as a status (201, 400, 401, 429, or 500) and JSON file. See Common Status and Error Codes [page 235] and Technical Constraints [page 288].

Response Example

201 "Success"

```
{
   "id":"484b6e1c-501c-4a07-85cb-84554656a175",
   "version":"2",
   "created":"2020-03-26T17:00:00.000000+00:00"
}
```

12.1.8.8 Activate Schema Version

Activate a particular version of a schema.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/schemas/<schemaId>/versions/<version>/activate`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`clientId`	Yes	String	*query*	The ID of the client used when creating the schema. Example: c_00
`schemaId`	Yes	String	*path*	The ID returned by the endpoint Create Schema [page 193]. Example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff
`version`	Yes	String	*path*	The `version` returned by the endpoint Create Schema Version [page 205]. Example: 2

Response

Response Fields

JSON Field	Description
`message`	Status message with information about the request

The response is given as a status (201, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235] and Technical Constraints [page 288].

Response Example

201 "Success"

```
{
    "message":"Schema version activated successfully."
}
```

12.1.8.9 Deactivate Schema Version

Deactivate a particular version of a schema..

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/schemas/`<schemaId>`/versions/`<version>`/deactivate`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client used when creating the schema. Example: c_00
schemaId	Yes	String	*path*	The ID returned by the endpoint Create Schema [page 193]. Example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff
version	Yes	String	*path*	The `version` returned by the endpoint Create Schema Version [page 205]. Example: 2

Response

Response Fields

JSON Field	Description
message	Status message with information about the request

The response is given as a status (201, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235] and Technical Constraints [page 288].

Response Example

201 "Success"

```
{
    "message":"Schema version deactivated successfully."
}
```

12.1.8.10 Add Fields to Schema Version

Add fields to schema version for a client.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/schemas/<schemaId>/versions/<version>/fields`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client used when creating the schema. Example: c_00

Parameter	Required	Data Type	Parameter Type	Description
`payload`	Yes	JSON Object	*body*	List containing the `headerFields` and `lineItemFields` that you want to add to the schema version.

You can also optionally use the `label` property to give user-friendly names to some or all of the `headerFields` and `lineItemFields` that you enter in the `payload`.

> → Remember
>
> - Each `label` can have a maximum length of 200 characters.
> - Each schema field `description` can have a maximum length of 500 characters.

Use setup types to choose the best schema field setup type for your documents. The following values are available:

- *auto*
- *manual*

In schemas created for standard document types, *auto* supports extraction in the following ways:

- Using the service's machine learning models: In this case, select an appropriate default extractor.
- Using generative AI: In this case, don't select a default extractor.

In schemas created for custom document types, *auto* supports extraction using generative AI. In this case, no default extractor is available.

> ⚠ Restriction
>
> The setup type *auto* is available without default extractor for schemas with the service plan **Document Information Extraction, premium edition** (`premium_edition`) only. See Service Plans [page 84] and Metering and Pricing [page 86].

Parameter	Required	Data Type	Parameter Type	Description
				⚠ **Caution** Always validate information extracted using generative AI before using it for critical applications. If you prefer not to use generative AI to extract information from documents, select the setup type *auto* with a default extractor (standard document types only) or select the setup type *manual* (standard and custom document types) when adding data fields to your schema. The setup type *manual* supports extraction using a template. It's available in schemas created for standard or custom document types. ⓘ **Note** To consume the setup types "*auto*" and "*manual*", use the `setupTypeVersion` 2.0.0. The `setupTypeVersion` 1.0.0 and the `setuptype` "*default*" are still supported. As of October 9, 2023, 2.0.0 is the recommended `setupTypeVersion`.
schemaId	Yes	String	*path*	The ID returned by the endpoint Create Schema [page 193]. Example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff
version	Yes	String	*path*	The `version` returned by the endpoint Create Schema Version [page 205]. Example: 2

Request Example: Payload with `label` and `setupTypeVersion` 2.0.0

```
{
    "headerFields":[
        {
            "name":"documentDate",
            "label":"Document Date",
            "description":"Document Date",
            "defaultExtractor":{

            },
            "setupType":"static",
            "setupTypeVersion":"2.0.0",
```

```
            "setup":{
                "type":"manual",
                "priority":1
            },
            "formattingType":"date",
            "formatting":{
                "dateformat":"dd/mm/yy"
            },
            "formattingTypeVersion":"1.0.0"
        }
    ],
    "lineItemFields":[
        {
            "name":"netAmount",
            "label":"Net Amount",
            "description":"Net Amount",
            "defaultExtractor":{
                "fieldName":"netAmount"
            },
            "setupType":"static",
            "setupTypeVersion":"2.0.0",
            "setup":{
                "type":"auto",
                "priority":1
            },
            "formattingType":"number",
            "formatting":{

            },
            "formattingTypeVersion":"1.0.0"
        },
        {
            "name":"discountAmount",
            "label":"Discount Amount",
            "description":"Discount Amount",
            "defaultExtractor":{

            },
            "setupType":"static",
            "setupTypeVersion":"2.0.0",
            "setup":{
                "type":"manual",
                "priority":1
            },
            "formattingType":"number",
            "formatting":{

            },
            "formattingTypeVersion":"1.0.0"
        }
    ]
}
```

Request Example: Payload with `label`**,** `setupType` *auto* **without** `defaultExtractor`**, and** `setupTypeVersion` **2.0.0**

```
{
    "headerFields":[
        {
            "name":"documentDate",
            "label":"Document Date",
            "description":"Document Date",
            "defaultExtractor":{

            },
            "setupType":"static",
            "setupTypeVersion":"2.0.0",
```

```
        "setup":{
            "type":"auto",
            "priority":1
        },
        "formattingType":"date",
        "formatting":{
            "dateformat":"dd/mm/yy"
        },
        "formattingTypeVersion":"1.0.0"
    },
    {
        "name":"documentNumber",
        "label":"Document Number",
        "description":"Document Number",
        "defaultExtractor":{

        },
        "setupType":"static",
        "setupTypeVersion":"2.0.0",
        "setup":{
            "type":"auto",
            "priority":1
        },
        "formattingType":"number",
        "formatting":{

        },
        "formattingTypeVersion":"1.0.0"
    }
],
"lineItemFields":[
    {
        "name":"netAmount",
        "label":"Net Amount",
        "description":"Net Amount",
        "defaultExtractor":{

        },
        "setupType":"static",
        "setupTypeVersion":"2.0.0",
        "setup":{
            "type":"auto",
            "priority":1
        },
        "formattingType":"number",
        "formatting":{

        },
        "formattingTypeVersion":"1.0.0"
    },
    {
        "name":"discountAmount",
        "label":"Discount Amount",
        "description":"Discount Amount",
        "defaultExtractor":{

        },
        "setupType":"static",
        "setupTypeVersion":"2.0.0",
        "setup":{
            "type":"auto",
            "priority":1
        },
        "formattingType":"number",
        "formatting":{

        },
        "formattingTypeVersion":"1.0.0"
```

```
        }
    ]
}
```

Request Example: Payload with `setupTypeVersion` **1.0.0**

```
{
    "headerFields":[
        {
            "name":"DocumentNumber",
            "description":"",
            "defaultExtractor":{
                "fieldName":"documentNumber"
            },
            "setupType":"static",
            "setupTypeVersion":"1.0.0",
            "setup":{

            },
            "formattingType":"string",
            "formatting":{

            },
            "formattingTypeVersion":"1.0.0"
        },
        {
            "name":"TaxId",
            "description":"",
            "defaultExtractor":{

            },
            "setupType":"static",
            "setupTypeVersion":"1.0.0",
            "setup":{

            },
            "formattingType":"string",
            "formatting":{

            },
            "formattingTypeVersion":"1.0.0"
        }
    ],
    "lineItemFields":[
        {
            "name":"Quantity",
            "description":"",
            "defaultExtractor":{
                "fieldName":"quantity"
            },
            "setupType":"static",
            "setupTypeVersion":"1.0.0",
            "setup":{

            },
            "formattingType":"number",
            "formatting":{

            },
            "formattingTypeVersion":"1.0.0"
        },
        {
            "name":"netAmount",
            "description":"",
            "defaultExtractor":{

            },
```

```
            "setupType":"static",
            "setupTypeVersion":"1.0.0",
            "setup":{

            },
            "formattingType":"number",
            "formatting":{

            },
            "formattingTypeVersion":"1.0.0"
        },
        {
            "name":"UnitPrice",
            "description":"",
            "defaultExtractor":{
                "fieldName":"unitPrice"
            },
            "setupType":"static",
            "setupTypeVersion":"1.0.0",
            "setup":{

            },
            "formattingType":"number",
            "formatting":{

            },
            "formattingTypeVersion":"1.0.0"
        }
    ]
}
```

Response

Response Fields

JSON Field	Description
message	Status message with information about the request

The response is given as a status (201, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235] and Technical Constraints [page 288].

Response Example

201 "Success"

```
{
    "message":"Schema fields have been uploaded successfully."
}
```

12.1.8.11 Get Schema Versions

Retrieve all versions for a schema.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/schemas/`<schemaId>`/versions`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client used when creating the schema. Example: c_00
schemaId	Yes	String	*path*	The ID returned by the endpoint Create Schema [page 193]. Example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff

Response

Response Fields

JSON Field	Description
created	Time when the schema was created
documentType	Type of the document submitted
documentTypeDescription	Description of the document submitted
id	ID of the schema
name	Name of the schema
predefined	True for standard documents, false for custom documents
schemaDescription	Description of the schema
state	State of the schema
updated	Time when the schema was updated
version	Version of the schema

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "schemas":[
        [
            {
                "name":"Basic Involve FormatSchema",
                "schemaDescription":"SAP Invoice Schema",
                "documentType":"Invoice",
                "documentTypeDescription":"Payment Advice with Order Number",
                "id":"484b6e1c-501c-4a07-85cb-84554656a175",
                "version":"1",
                "predefined":"True",
                "created":"2020-03-26T17:00:00.000000+00:00",
                "updated":"2020-04-26T17:00:00.000000+00:00",
                "state":"draft"
            },
            {
                "name":"Basic Involve FormatSchema",
                "schemaDescription":"SAP Invoice Schema",
                "documentType":"Invoice",
                "documentTypeDescription":"Payment Advice with Order Number",
                "id":"484b6e1c-501c-4a07-85cb-84554656a175",
                "version":"2",
                "predefined":"True",
                "created":"2020-03-26T17:00:00.000000+00:00",
                "updated":"2020-04-26T17:00:00.000000+00:00",
                "state":"draft"
            }
        ]
    ]
}
```

12.1.8.12 Get Schema Version Details

Retrieve version details of a schema for a client.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: /document-information-extraction/v1

URL Endpoint Path: /schemas/<schemaId>/versions/<version>

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client used when creating the schema. Example: c_00

Parameter	Required	Data Type	Parameter Type	Description
schemaId	Yes	String	*path*	The ID returned by the endpoint Create Schema [page 193]. Example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff
version	Yes	String	*path*	The version returned by the endpoint Create Schema Version [page 205]. Example: 2

Response

Response Fields

JSON Field	Description
created	Time when the schema was created
documentType	Type of the document submitted
documentTypeDescription	Description of the document submitted
id	ID of the schema
name	Name of the schema
predefined	True for standard documents, false for custom documents
schemaDescription	Description of the schema
state	State of the schema
updated	Time when the schema was updated

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "schemas":[
        [
            {
                "name":"Basic Involve FormatSchema",
                "schemaDescription":"SAP Invoice Schema",
                "documentType":"Invoice",
                "documentTypeDescription":"Payment Advice with Order Number",
                "id":"484b6e1c-501c-4a07-85cb-84554656a175",
                "predefined":"True",
                "created":"2020-03-26T17:00:00.000000+00:00",
                "updated":"2020-04-26T17:00:00.000000+00:00",
                "state":"draft"
            },
            {
                "name":"Daimier Payment Advice Schema",
                "schemaDescription":"Payment Advice Schema",
```

```
            "documentType":"Payment Advice",
            "documentTypeDescription":"Payment Advice with Order Number",
            "id":"484b6e1c-501c-4a07-85cb-84554656a189",
            "predefined":"False",
            "created":"2020-03-26T17:00:00.000000+00:00",
            "updated":"2020-04-26T17:00:00.000000+00:00",
            "state":"active"
        }
    ]
  ]
}
```

12.1.8.13 Delete Schema Versions

Delete versions associated with a schema.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/schemas/`<schemaId>`/versions`

HTTP Method: *DELETE*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client used when creating the schema. Example: c_00
payload	Yes	JSON Object	*body*	Comma-separated list of the schema versions you want to delete. The schema and all its versions are deleted if `payload` is empty. You can't delete version "1".
schemaId	Yes	String	*path*	The ID returned by the endpoint Create Schema [page 193]. Example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff

Request Example

```
{
    "version":[
        "5"
    ]
}
```

Response

Response Fields

JSON Field	Description
message	Status message with information about the request

The response is given as a status (201, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

201 "Success"

```
{
    "message":"Schema versions deleted successfully."
}
```

12.1.9 Template API

Create, reuse, edit, and delete templates based on schemas and document types. You can select templates together with a corresponding schema to extract information from business documents of the appropriate type and structure. The Template API provides endpoints to create, update, list, import, export, activate, deactivate, and delete templates. You can also associate documents with a template and dissociate documents from a template using the Template API endpoints.

The Template API consists of the following endpoints:

- Create or Update Template [page 220]
- Get Template [page 221]
- Import Template [page 224]
- Get Template Details [page 224]
- Delete Template [page 226]
- Activate Template [page 227]
- Deactivate Template [page 228]
- Associate Document with Template [page 229]
- Dissociate Document from Template [page 230]
- Export Template [page 231]
- Create Template Metadata [page 232]
- Get Template Metadata [page 233]

12.1.9.1 Create or Update Template

Create or update a template.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/templates`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
`payload`	Yes	JSON Object	*body*	List containing `id`, `name`, `description`, `clientId`, `schemaId`, and `schemaVersion`.
				ⓘ Note
				If `id` is not provided, a template ID is generated and returned.
				If `id` is provided, but it is not in the system, a new record with template ID provided is created.
				If `id` is provided, and it is in the system, an update of record with same template ID is executed.

Request Example

```
{
    "id":"37c8a59b-b210-48c1-9002-19ec989066eb",
    "name":"Test_Template",
    "description":"Test description",
    "clientId":"c_00",
    "schemaId":"37c8a59b-b210-48c1-9002-19ec989066eb",
    "schemaVersion":"1"
}
```

Response

Response Fields

JSON Field	Description
id	Template ID

The response is given as a status (201, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235] and Technical Constraints [page 288].

Response Example

201 "Created"

```
{
    "id":"31516520-b4c9-40a6-b9ba-94d1800d472d"
}
```

12.1.9.2 Get Template

Get templates for a `client` ID.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/templates`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. Example: c_00
includeHeader	No	Boolean	*query*	Result should include header fields
includeLineItems	No	Boolean	*query*	Result should include line item fields
limit	No	Integer	*query*	Max number of records to be returned. All records are returned if limit = 0. Example: 10

Parameter	Required	Data Type	Parameter Type	Description
offset	No	Integer	query	The offset of the query result start index to be returned. Example: 0
order	No	String	query	Order the results. Example: name asc
schemaId	No	String	query	The ID returned by the endpoint Create Schema [page 193]. Example: 4476cc01-72f3-4b64-9eb0-cdd9c1cb27ff

Response

Response Fields

JSON Field	Description
clientId	ID of the client this template was created for
creationDate	Date when you created this template
documentAssociations	ID of the documents associated with this template
description	Template description
documentType	Type of the document this template was created for
extraction	Dictionary containing all the extracted header fields and line items
headerFields	Dictionary containing all extracted header fields
id	Template ID
isActive	Set to true if template has been activated. Set to false if template has not been activated, or it has been deactivated
language	Template language
lastUpdatedDate	Date when you last updated this template
lineItemFields	Dictionary containing all extracted line items
name	Template name
results	List containing information of all templates by clientId
schemaId	Schema ID
status	Template status. Possible values: "NO_SAMPLES", "NO_ANNOTATIONS", or "READY"

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "results":[
```

```
    {
        "id":"5fb6279a-1bb9-4e37-b3bc-95ffb0e3d220",
        "schemaId":"3e048fac-7799-45dc-a360-ff921d8ef152",
        "name":"Test Template",
        "description":"Test Description",
        "language":"en",
        "documentType":"invoice",
        "clientId":"c_00",
        "status":"NO_SAMPLES",
        "isActive":true,
        "creationDate":"2023-11-14T07:39:23.536547+00:00",
        "lastUpdatedDate":"2023-11-14T07:39:23.536547+00:00",
        "documentAssociations":[
            {
                "id":"sample_id"
            }
        ],
        "extraction":{
            "headerFields":[
                {
                    "name":"documentNumber",
                    "label":"Document Number:",
                    "type":"number"
                }
            ]
        }
    },
    {
        "id":"1213723c-bdff-4b2a-b821-93f051966b0c",
        "schemaId":"0f68b9c8-1e10-467d-a01a-23ffae9b5e4e",
        "name":"Test Template 2",
        "description":"Test Description 2",
        "language":"en",
        "documentType":"invoice",
        "clientId":"c_00",
        "status":"NO_SAMPLES",
        "isActive":false,
        "creationDate":"2023-11-14T07:39:23.536547+00:00",
        "documentAssociations":[
            {
                "id":"sample_id"
            }
        ],
        "extraction":{
            "headerFields":[
                {
                    "name":"documentNumber",
                    "type":"number"
                }
            ]
        }
    }
  ]
}
```

12.1.9.3 Import Template

Create or update a template.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/templates/import`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. Example: c_00
file	Yes	File	*formData*	The template file you want to import.

Response

The response is given as a status (201, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

201 "Created"

12.1.9.4 Get Template Details

Get template details for a template ID. You can only get template details that belong to the same `zone_id` and `client_id`.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/templates/<template_id>`

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. Example: c_00
template_id	Yes	String	*path*	Template ID. Example: 4176cc01-71f3-4b64-9eb0-cdd9c0cb27fd

Response

Response Fields

JSON Field	Description
clientId	ID of the client this template was created for
creationDate	Date when you created this template
documentAssociations	ID of the documents associated with this template
description	Template description
documentType	Type of the document this template was created for
extraction	Dictionary containing all the extracted header fields and line items
headerFields	Dictionary containing all extracted header fields
id	Template ID
isActive	Set to true if template has been activated. Set to false if template has not been activated, or it has been deactivated
language	Template language
lastUpdatedDate	Date when you last updated this template
lineItemFields	Dictionary containing all extracted line items
name	Template name
schemaId	Schema ID
schemaName	Schema Name
status	Template status. Possible values: "NO_SAMPLES", "NO_ANNOTATIONS", or "READY"

The response is given as a status (200, 400, 401, or 500) and JSON file. See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "id":"37c8a59b-b210-48c1-9002-19ec989066eb",
```

```
        "schemaId":"608aa59c-4895-4308-bcae-905f8f343acc",
        "name":"Test Template",
        "description":"Test Template Description",
        "language":"en",
        "documentType":"invoice",
        "clientId":"c_00",
        "status":"NO_SAMPLES",
        "isActive":true,
        "creationDate":"2023-11-14",
        "lastUpdatedDate":"2023-11-14T07:39:23.536547+00:00",
        "schemaName":"SAP_Schema",
        "documentAssociations":[
            {
                "id":"f58f7e0b-a1a8-449c-aa4b-6c71e256cd3e"
            }
        ],
        "extraction":{
            "headerFields":[
                {
                    "name":"string",
                    "label":"string",
                    "type":"string"
                }
            ],
            "lineItemFields":[
                {
                    "name":"string",
                    "label":"string",
                    "type":"string"
                }
            ]
        }
    }
}
```

12.1.9.5 Delete Template

Delete a template and its links to the associated documents for a template ID.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: /document-information-extraction/v1

URL Endpoint Path: /templates/<template_id>

HTTP Method: *DELETE*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. Example: c_00

Parameter	Required	Data Type	Parameter Type	Description
template_id	Yes	String	*path*	Template ID. Example: 4176cc01-71f3-4b64-9eb0-cdd9c0cb27fd

Response

Response Fields

JSON Field	Description
message	Status message with information about the request

The response is given as a status (200, 400, 401, or 500). See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "message":"Successfully deleted 1 template."
}
```

12.1.9.6 Activate Template

Activate a template.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: /document-information-extraction/v1

URL Endpoint Path: /templates/<template_id>/activate

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. Example: c_00
template_id	Yes	String	*path*	Template ID. Example: 4176cc01-71f3-4b64-9eb0-cdd9c0cb27fd

Response

Response Fields

JSON Field	Description
message	Status message with information about the request

The response is given as a status (200, 400, 401, or 500). See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "message":"Successfully activated the template"
}
```

12.1.9.7 Deactivate Template

Deactivate a template.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: /document-information-extraction/v1

URL Endpoint Path: /templates/<template_id>/deactivate

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. Example: c_00
template_id	Yes	String	*path*	Template ID. Example: 4176cc01-71f3-4b64-9eb0-cdd9c0cb27fd

Response

Response Fields

JSON Field	Description
message	Status message with information about the request

The response is given as a status (200, 400, 401, or 500). See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "message":"Successfully deactivated the template"
}
```

12.1.9.8 Associate Document with Template

Associate a document with a template.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: /document-information-extraction/v1

URL Endpoint Path: /templates/<template_id>/documents/<document_id>

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. Example: c_00
document_id	Yes	String	*path*	Document ID. Example: 5146ce05-7tf3-4g64-9eb0-vaa9f9cb22af
template_id	Yes	String	*path*	Template ID. Example: 4176cc01-71f3-4b64-9eb0-cdd9c0cb27fd
validateDocu mentStatus	No	Boolean	*query*	Set to false to skip document status validation when associating document with template. The default value is true.

Response

Response Fields

JSON Field	Description
message	Status message with information about the request

The response is given as a status (200, 400, 401, or 500). See Common Status and Error Codes [page 235].

Response Example

200 "Success"

```
{
    "message":"Successfully added document to the template."
}
```

12.1.9.9 Dissociate Document from Template

Dissociate a document from a template.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: /document-information-extraction/v1

URL Endpoint Path: /templates/<template_id>/documents/<document_id>

HTTP Method: *DELETE*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. Example: c_00
document_id	Yes	String	*path*	Document ID. Example: 5146ce05-7tf3-4g64-9eb0-vaa9f9cb22af
template_id	Yes	String	*path*	Template ID. Example: 4176cc01-71f3-4b64-9eb0-cdd9c0cb27fd

Response

Response Fields

JSON Field	Description
message	Status message with information about the request

The response is given as a status (200, 400, 401, or 500). See .

Response Example

200 "Success"

```
{
    "message":"Successfully removed document from the template."
}
```

12.1.9.10 Export Template

Export a template.

> ⓘ Note
>
> You can download malware-scanned documents only. You can't download documents that are part of the template export package but haven't been malware-scanned during upload.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: /document-information-extraction/v1

URL Endpoint Path: /templates/<template_id>/export

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. Example: c_00
template_id	Yes	String	*path*	Template ID. Example: 4176cc01-71f3-4b64-9eb0-cdd9c0cb27fd

Response

The response is given as a status (200, 400, 401, 410 or 500). See Common Status and Error Codes [page 235].

Response Example

200 "Success"

12.1.9.11 Create Template Metadata

Set certain fields of a template to be fixed-value fields.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: `/document-information-extraction/v1`

URL Endpoint Path: `/templates/`<`template_id`>`/metadata`

HTTP Method: *POST*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. Example: c_00
payload	Yes	JSON Object	*body*	List containing all fixed-value fields of a template. You can use the options `extraction` and `isLineItemField` to provide extraction model information for a field. Possible values for `extraction` are **template**, or **schema**, and for `isLineItemField` are **true**, or **false**.
template_id	Yes	String	*path*	Template ID. Example: 4176cc01-71f3-4b64-9eb0-cdd9c0cb27fd

Request Payload Examples

```
{
    "metadata":[
        {
            "name":"name",
            "value":"value"
        }
    ]
```

```
        }
```

```
{
    "metadata":[
        {
            "name":"name",
            "value":"value",
            "extraction":"template",
            "isLineItemField":"false"
        },
        {
            "name":"name",
            "value":null,
            "extraction":"schema",
            "isLineItemField":"true"
        }
    ]
}
```

Response

Response Fields

JSON Field	Description
id	Metadata ID

The response is given as a status (201, 400, 401, or 500). See Common Status and Error Codes [page 235].

Response Example

201 "Accepted"

```
{
    "id":"b6e6ddaf-ceb0-4245-ab07-6ced50b18807"
}
```

12.1.9.12 Get Template Metadata

Get all fixed-value fields of a template.

Request

Base URL: url value from outside the uaa section of the service key

URL Path Extension: /document-information-extraction/v1

URL Endpoint Path: /templates/<template_id>/metadata

HTTP Method: *GET*

Request Parameters

Parameter	Required	Data Type	Parameter Type	Description
clientId	Yes	String	*query*	The ID of the client. Example: c_00
template_id	Yes	String	*path*	Template ID. Example: 4176cc01-71f3-4b64-9eb0-cdd9c0cb27fd

Response

Response Fields

JSON Field	Description
extraction	Either **template** or **schema**
isLineItemField	Either **true** or **false**
metadata	Dictionary containing all fixed-value fields of a template
name	Metadata name
value	Metadata value

The response is given as a status (200, 400, 401, or 500). See Common Status and Error Codes [page 235].

Response Examples

200 "Success"

```
{
    "metadata":[
        {
            "name":"name",
            "value":"value"
        }
    ]
}
```

200 "Success"

```
{
    "metadata":[
        {
            "name":"name",
            "value":"value",
            "extraction":"template",
            "isLineItemField":"false"
        },
        {
            "name":"name",
            "value":null,
            "extraction":"schema",
            "isLineItemField":"true"
        }
    ]
}
```

12.1.10 Common Request Headers

Header	Required	Description
Authorization	Yes	Access token used to access the service.
Content-Type	Yes	Indicates the type that can be consumed by this service. Set the `Accept` parameter to **application/json**.
Accept	Yes	Indicates the type that is associated with the message body. Set the `Content-Type` parameter to **application/json**.
tenantName	Yes	Specifies the tenant name used to access this service.

12.1.11 Common Status and Error Codes

Code	Reason
200	The request was successful
201	Client creation, document upload and deletion of enrichment data or uploaded document, for example, was successful.
400	Bad request. Document Information Extraction process could not be submitted or completed; possibly due to parameter error.
401	Unauthorized, for example, no token or bad token
413	The request you are making is too large. Either you are sending a file that is too large or trying to process too many objects in a single request. See Technical Constraints [page 288].
415	Unsupported document file format. See Supported Document Types and File Formats [page 91].

Code	Reason
422	Unprocessable entity. Your request payload references a clientId, senderName, or documentNumber that does not exist. For example, you will get this error if you try to create a document for a client that does not exist. You may also get this error if the document you upload cannot be parsed.
429	Application quota limit exceeded
500	Internal server error. Document Information Extraction process could not be submitted or completed; possibly due to an internal error.
503	System is temporarily unavailable

12.2 Notifications

Discover the notifications functionality.

Use this functionality to get notifications about your processed documents status without having to constantly poll the Document Information Extraction service. Through this functionality, Document Information Extraction notifies an endpoint using a callback URL that you specify with the explicit name of `document-information-extraction-callback`. The notification callback request will only be sent once document processing has either completed or failed.

> ⓘ Note
>
> To enable the notifications functionality, set the configuration key `activateDocumentNotifications` to *true* as described in Create Configuration [page 122].
>
> Document Information Extraction sends only one notification per document without retry.

> ⚠ Restriction
>
> The notifications functionality is available from 2020-05-18. Any service instance created before this date does not include this functionality. If existing customers want to use their existing instances with this new functionality, they need to subscribe to the Document Information Extraction UI in SAP Business Technology Platform, as described in Subscribing to the Document Information Extraction UI [page 243] (procedure steps from 1 until 6).

Related Information

Enabling Destination Service for Notifications [page 237]
Creating Destination Configuration for Notifications [page 238]

12.2.1 Enabling Destination Service for Notifications

Prerequisites

You have subscribed to the Document Information Extraction UI in SAP Business Technology Platform.

> → Tip
>
> In Subscribing to the Document Information Extraction UI [page 243], observe the prerequisites and follow the procedure steps from 1 until 4.

To use the notifications functionality, you need to enable the Cloud Foundry Destination Service at subaccount level via the *Entitlements*. After that, *Destinations* will be visible in the left navigation pane.
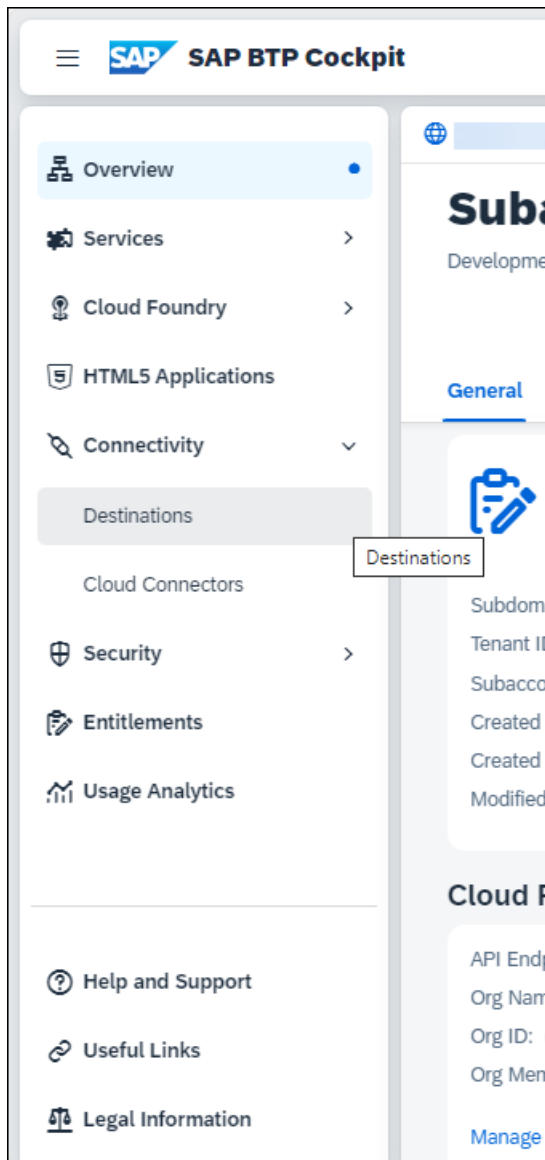
See Consuming the Destination Service.

12.2.2 Creating Destination Configuration for Notifications

Create a new destination configuration that includes the callback URL, and some additional information about authentication credentials and the ProxyType.

Name the callback endpoint `document-information-extraction-callback`. You can only have one callback endpoint with this name on subaccount level. This destination configuration callback URL must link to an endpoint connected to the Internet.
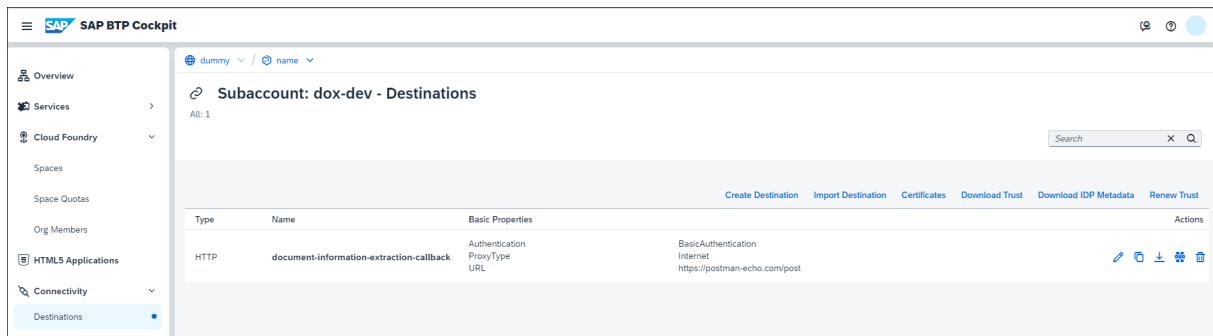
See Create HTTP Destinations.

Example

NoAuthentication Destination Configuration:

Example

BasicAuthentication Destination Configuration:

Example

OAuth2 Client Credentials Destination Configuration:

12.2.3 Supported Authentication Methods

The following authentication types are currently supported:

- NoAuthentication
- BasicAuthentication
- OAuth2 Client Credentials

See Create HTTP Destinations.

12.2.4 Callback Request Examples

The Document Information Extraction callback sends a POST request to the URL specified in the destination configuration with the name `document-information-extraction-callback`.

Example

Payload

The payload will be sent with the POST request to the specified callback URL in the destination configuration specified by the customer.

The payload includes the ID of the uploaded document and its status. These two fields are in alignment with the other Document Information Extraction API fields:

- The ID field string represents a `<uuid>` representing a document
- The status field string includes the process status which can be either "DONE", or "FAILED"

This payload indicates a successful processing of the document:

```
{
        "id": "d7c08124-d852-408f-8d46-0466312f6007",
        "status": "DONE"
}
```

Example

NoAuthentication

CURL representation of the POST request with no authentication to the callback URL of the customer:

```
curl --location --request POST 'https://callback.url/notify-callback' \
--header 'Content-Type: application/json' \
--data-raw '{
    "id": "d7c08124-d852-408f-8d46-0466312f6007",
    "status": "DONE"
}'
```

Example

BasicAuthentication

CURL representation of the POST request with basic authentication to the callback URL of the customer:

```
curl --location --request POST 'https://callback.url/notify-callback' \
--header 'Content-Type: application/json' \
--header 'Authorization: Basic dGVzdC11c2VyOnRlc3QtcGFzc3dvcmQ=' \
--data-raw '{
    "id": "d7c08124-d852-408f-8d46-0466312f6007",
    "status": "FAILED"
}'
```

Example

OAuth2 Client Credentials

CURL representation of the POST request with OAuth2 client credentials to the callback URL of the customer:

```
curl --location --request POST 'https://callback.url/notify-callback' \
--header 'Content-Type: application/json' \
--header 'Authorization: Bearer
eyJhbGciOiJSUzI1NiIsImprdSI6Imh0dHBzOi8vc2FwLXByb3Zpc2lvbmluZy5hdXRoZW50aWNhdGlvbi5zYXVaGFuY5vbmRlWFuC5jb20vdG9rZW5fa2V5cyIsImtpZCI6ImtleS1pZC0xIiwidHlwIjoiSldUIn0.eyJqdGkiOiIxNzBhYzJjU0YmQwOTE0NDhkNjBhZDcyMDQzNyIsImV4dF9hdHRyIjp7ImVuaGFuY2VyIjoiWFNVQUEiLCJ6ZG4iOiJzYXAtcHJvdmlzaW9uaW5nI6InNiLXRlbmFudC1vbmJvYXJkaW5nIXQ
```

```
xMyIsInNjb3BlIjpbImRveC14c3VhYS1pbnQtdCFiOTM4MC5DYWxsYmFjayJdLCJjbGllbnRfaWQiOiJz
Yi10ZW5hbnQtb25ib2FyZGluZyF0MTIiLCJjaWQiOiJzYi10ZW5hbnQtb25ib2FyZGluZyF0MTIiLCJhe
nAiOiJzYi10ZW5hbnQtb25ib2FyZGluZyF0MTIiLCJncmFudF90eXBlIjoiY2xpZW50X2NyZWRlbnRpYW
xzIiwicmV2X3NpZyI6Ijc3MWQ1DDFmIiwiaWF0Njk4LCJleHAiOjE1ODUxNzQ4OTgsImlzcyI6Imh0dHA
6Ly9zYXAtcHJvdmlzaW9uaW5nLmxvY2FsaG9zdDo4MDgwL3VhYS9vYXV0aC90b2tlbiIsInpppZCI6InNh
cC1wcm92aXNpb25pbmciLCJhdWQiOlsic2ItdGVuYW50LW9uYm9hcmRpbmchdDEzIiwiZG94LXhzdWFhL
WludC10IWI5MzgwIl19.ROCb2LQZOGTFE7ZKQVC8T-
kuvzb8DtMjetY8vqeJUt9GC1UA24siGkiagTGPYNzalvlBwLW2b1Thx7WA3OkVIMLiWwG_7AHm6ONjoUz
Ew8v35NMlHALrY97oRPgSZOSCWFzhzKnL6t1Y0G0m83ctQAJaml-wd5NdDSbHyoIkJ3i5qhXC-
rVaNsAnfX9eerJtjYwxvqvIYi9rEewTg-EcRBdWndvB962RFDGDZco_92ZNP4uYN238_0-
ylFKYFF8mdlSivwc8SNscXCojlCAgk_4kYqiM_3ai5FkuXwyZunoPtrNnr77yK5HUyuZUuYmhzy7F6GJI
59VCrPYnELJPiw' \
--data-raw '{
    "id": "d7c08124-d852-408f-8d46-0466312f6007",
    "status": "DONE"
}'
```

12.2.5 Callback Response Status

The status of the callback response should be 200 "OK", as you can see in the curl response below. Statuses less than 400 are also accepted.

Request

Callback request from the Document Information Extraction service:

```
curl --verbose --location --request POST 'https://callback.url/notify-callback'
\
--header 'Authorization: Basic dGVzdC11c2VyOnRlc3QtcGFzc3dvcmQ=' \
--header 'Content-Type: application/json' \
--data-raw '{
"id": "d7c08124-d852-408f-8d46-0466312f6007",
"status": "DONE"
}'
```

Response

```
< HTTP/1.1 200 OK
< Content-Type: application/json; charset=utf-8
< Date: Thu, 16 Apr 2020 06:55:41 GMT
<
{}
```

> ⓘ Note
>
> The body of the callback response is not relevant to the Document Information Extraction service, only the response status of 200.

13 Using the Document Information Extraction UI

Find out how to subscribe to, access, and use the Document Information Extraction UI.

Related Information

13.1 Subscribing to the Document Information Extraction UI

To use the Document Information Extraction UI and other features, you need to subscribe to the service UI in SAP Business Technology Platform (SAP BTP).

Prerequisites

- You have an SAP BTP global account and a Cloud Foundry subaccount.
- You're a global account administrator.
- You've created a service instance for Document Information Extraction.
- You've created business users and user groups in your identity provider (IdP). SAP ID Service is the default IdP, but you can also add your instance of the Identity Authentication service or a different IdP.

> ⓘ Note
>
> If you use the Identity Authentication service, see Establish Trust and Federation Between UAA and Identity Authentication.
>
> If you use a different IdP, see Establish Trust and Federation with UAA Using Any SAML Identity Provider.

Context

> **→ Tip**
>
> You can also use the *Set up account for Document Information Extraction* booster in the SAP BTP cockpit to automate the process. In this case, you don't need to perform the steps for subscribing to the Document Information Extraction UI described here. See Boosters and the tutorial Use Free Tier to Set Up Account for Document Information Extraction and Go to Application📌.

> **ⓘ Note**
>
> You can create multiple service instances for Document Information Extraction. However, we recommend creating only one, unless there's a compelling reason for having more.
>
> If you do use more than one instance, you can change between instances by choosing *Settings* (⚙ *cogwheels icon*) *Change Instance* on the Document Information Extraction UI. You can specify the instance by entering its name or its ID.

To subscribe to the Document Information Extraction UI, do the following.

Procedure

1. Open the SAP BTP cockpit and go to your subaccount.
2. Click *Service Marketplace* under *Services* on the left navigation pane.
3. Search for Document Information Extraction and click the tile.

   The *Overview* page appears.
4. Click *Create*.

   The *New Instance or Subscription* dialog appears.

   > **→ Remember**
   >
   > Before proceeding, check whether you've created an instance for Document Information Extraction. If you haven't, create the service instance before continuing with the following steps. Creating a service instance is a prerequisite for using the Document Information Extraction UI.

5. Choose the *default Subscription* plan.
6. Click *Create*.
7. Click *Users* under *Security* on the left navigation pane.
8. Click the arrow under *Actions* in the row with your user.

   The *Overview* page appears.
9. Click *Assign Role Collection*.
10. Select the role collection that you wish to assign. See Role Collections [page 245].
11. Click *Assign Role Collection*. For more information, see Assign Users to Role Collections.
12. Click *Instances and Subscriptions* on the left navigation pane.

13. Click the three dots at the right end of the row with the Document Information Extraction application and select *Go to Application* from the dropdown.

The logon screen appears.

14. Enter your *User* and *Password*, previously created in your identity provider (IdP), to log on.

> ⓘ Note
>
> You may not have to log on explicitly at this point if the following is true:
> - You've configured your user to log in with a certificate.
> - Your user already has an active session on your IdP.

The Document Information Extraction UI appears.

13.1.1 Role Collections

Find out about the role collections you can use with the Document Information Extraction UI.

Document Information Extraction provides default role collections that you can assign to users. These role collections determine which actions a user can carry out on the Document Information Extraction UI.

The default role collections and associated actions are as follows:

Role Collection	Actions
`Document_Information_Extraction_UI_Templates_Admin`	Manage the template and schema lifecycle. View documents and edit extraction results.
`Document_Information_Extraction_UI_End_User`	View documents, edit extraction results, and work with schemas and templates.
`Document_Information_Extraction_UI_Document_Viewer`	View documents in the UI application.

The default role collections grant users the following read/write permissions:

	Document		Template/Schema	
	Read	Write	Read	Write
`Document_Information_Extraction_UI_Document_Viewer`	✅			
`Document_Information_Extraction_UI_End_User`	✅	✅	✅	

```
Document_Inform     ✓              ✓              ✓              ✓
ation_Extractio
n_UI_Templates_
Admin
```

> → Remember
>
> The role collection `Document_Information_Extraction_UI_Admin_User` has been deprecated.
> Assign the `Document_Information_Extraction_UI_Templates_Admin` to any administrators who
> formerly used the deprecated role collection to manage the template and schema lifecycle.

13.2 Using the Key Features of the Document Information Extraction UI

Find out how to use the Document Information Extraction UI features for documents, schemas, and templates.

Use the following features to handle a wide range of tasks:

- Document [page 250]
- Schema Configuration [page 256]
- Template [page 266]

> ⓘ Note
>
> For recommendations on getting better extraction results, see Optical Character Recognition (OCR): Best
> Practices [page 272].

For instructions on how to set the language of the Document Information Extraction UI, see Set Screen
Language [page 248].

For information about how to use the integrated digital assistant to find answers to support-related questions,
see Built-In Support [page 249].

13.2.1 UI Settings

Find out about the UI settings for the Document Information Extraction UI.

You can change UI settings by choosing the dropdown for your user name at the top right of the screen and
then choosing ▷ *User Settings* 〉 *UI Settings* 〉.

Here, you can set the thresholds for the extraction confidence ranges (high, medium, and low) and the color
associated with each range.

You can also enable features that aren't available by default on the Document Information Extraction UI.

Related Information

13.2.1.1 Enable Features

Find out how to enable features that aren't available by default on the Document Information Extraction UI.

Context

Some features of the Document Information Extraction UI aren't available by default. If you wish to use these features, you must first enable them.

To do so, proceed as follows:

Procedure

1. Open the dropdown for your user name at the top right of the screen.
2. Select *User Settings* (⚙ , *settings*).
3. Select *UI Settings* (⚙ , *customize*)
4. Select *Feature Enablement*.

   > ⓘ Note
   >
   > Currently, *Schema Field Categories* is the only feature available here.

5. Set *Schema Field Categories* to *On* if you wish to use field categories in addition to the standard features.

   > → Remember
   >
   > This setting is browser specific: If you enable the feature in one browser and later open the Document Information Extraction UI in another, you must repeat the preceding steps before you can use the feature.

13.2.1.2 Change Confidence Range Settings

Find out how you can customize confidence ranges and the associated color coding on the Document Information Extraction UI.

Context

When you use the *Document* feature to view the extraction results for documents uploaded to the Document Information Extraction service, you see the extraction confidence ranges for header fields and line items.

The default colors for these ranges are green for high, orange for medium, and red for low.

By default, the lower thresholds for the confidence ranges are as follows:

- High: 80%
- Medium: 51%
- Low: 0%

You can change the default color and the lower threshold of each confidence range by doing the following:

Procedure

1. Open the dropdown for your user name at the top right of the screen.
2. Select *User Settings* (⚙, *settings*).
3. Select *UI Settings* (⚙, *customize*)
4. Choose *Confidence Level*.
5. Set the lower threshold for one or more of the confidence ranges *High*, *Medium*, and *Low*, as desired.
6. Set the color for one or more of the confidence ranges, as desired.
7. Choose *Apply* when you've completed your changes.

   If you wish to revert to the default settings, choose *Reset*.

13.2.2 Set Screen Language

Select the screen language for the Document Information Extraction UI.

Context

The Document Information Extraction UI is currently available in the following languages:

Language	Language Code
German	de
English	en
Spanish	es
French	fr
Italian	it
Japanese	ja
Korean	ko
Portuguese	pt
Russian	ru
Chinese Simplified	zh_CN
Chinese Traditional	zh_TW

> ⓘ **Note**
>
> The SAP Companion in-app help is also available in the language that you select for the UI. Display this help by choosing ⓘ *(question mark)* in the top right of the screen.

To set the screen language, do the following:

Procedure

1. Open the dropdown for your user name at the top right of the screen.
2. Select *Languages*.
3. Select your preferred language.
4. Complete your entries by choosing *Apply*.

13.2.3 Built-In Support

Use the integrated digital assistant on the Document Information Extraction UI to quickly find answers to your support-related questions.

Context

The Document Information Extraction UI includes Built-In Support, an embedded digital assistant that allows you to search for support-related information without leaving the UI.

> ⓘ **Note**
>
> If you have an s-user ID and the associated authorizations, Built-In Support also allows you to report issues, review cases, and chat with an expert or a chatbot.

Procedure

1. Choose *Built-In Support* (🎧 *headset icon*).

   The Built-In Support initial screen appears. This screen gives you access to the basic support functions that are available to all users. Here, you can enter keywords in the intelligent search field to find relevant information in the documentation for Document Information Extraction UI. You can also call up recommended information about the service directly via the links provided.

2. Choose the *Help Information* (ⓘ *hint icon*).

   The *Contextual Help* screen appears. Here, you can access information, including tutorial videos, the Built-In Support documentation, the privacy statement, and the terms of use.

3. Choose 🗛 *(person icon)* to view system context information.

   If you have an s-user ID, you can sign in to access more Built-In Support functions. These functions allow you to report issues via case or by chatting with an expert. In addition, you can review your cases.

13.2.4 Document

Use this Document Information Extraction UI feature to upload documents to the service and get machine learning predictions for the extracted header fields and line items.

Context

Use this feature to do the following:

- Add Document [page 251]
- View and Edit Extraction Results [page 252]
- Delete Documents [page 255]

For additional information on working with documents, see the best practices under Document: Best Practices [page 283].

13.2.4.1 Add Document

Procedure

1. Open the Document Information Extraction UI, as described in Subscribing to the Document Information Extraction UI [page 243].
2. Click the *Document* icon in the left navigation pane.
3. Click *Upload a new document* (✚ *add icon*) at the top right of the screen.

   The *Select Document* area appears. Here, you can upload a maximum of 50 files. Add files individually or select a folder containing multiple files. Each file can have a maximum size of 50 MB and 100 pages. The service supports the following document types: invoice, payment advice, purchase order, and custom in PDF, JPG, PNG, and TIFF format.
4. Select the document type.
5. Choose a schema. This step is mandatory.

   > ⓘ Note
   >
   > The Document Information Extraction UI automatically proposes an SAP schema that matches the selected document type. Preconfigured SAP schemas are available for the following standard document types: purchase order, payment advice, and invoice. In addition, there's an SAP schema for custom documents (`SAP_OCROnly_schema`).
   >
   > If your administrator has created and activated other schemas for the selected document type, you can select one from the *Schema* dropdown.

6. Optionally choose a template, making sure that it matches the document type and schema that you selected. You can also use the *Detect automatically* function to get the service to search for the correct template.

   > → Remember
   >
   > Templates are available only if your administrator has created and activated them.

7. Upload one or more document files by dragging and dropping them or by clicking ✚ *(add icon)*.
8. Click *Confirm*

   You now see the documents that you've uploaded, with *Document Name*, *Upload Date*, and *Status*. When the header fields and line items defined in the selected schema have been extracted, the document status changes from "PENDING" to "READY". You can now review the extraction results and make any corrections. If an error occurs during document processing, the status changes from "PENDING" to "FAILED". In this case, you must upload the document again.
9. In the top right of the screen, you see the `clientId` (c_00, for example) of the listed uploaded documents. Click *Change Client* and select another `clientId` (c_01, for example) to see the list of uploaded documents that have a different `clientId`.

   Before you can change clients, there must be at least one client in addition to *Default*. You can't create clients on the Document Information Extraction UI. To add new clients, use `Swagger UI` and follow the steps in Create Client [page 114].

> ⓘ Note
>
> You can restrict user access to specified clients by using the `clientSegregation` configuration key. For more details and guidance, see Configuration Keys [page 124] and Client Segregation in Document Information Extraction: A Brief Guide↪.

13.2.4.1.1 Download Troubleshooting Data

Find out how to download data needed to troubleshoot issues with adding documents to the Document Information Extraction UI.

Context

For each document that you add to the Document Information Extraction UI, you can download a zip folder with files for troubleshooting.

Procedure

1. Choose the *Document* icon in the navigation on the left of the screen.
2. Now, choose a document to display its details.

   The details pane appears on the right of the screen.

3. Choose 🔍 *(Download Troubleshooting Data)* to the right of the document status at the top of the details pane.

   The Document Information Extraction UI downloads a zip folder to your local machine. The files in the folder include the document that you uploaded as well as details of the document, template, and schema.

   You can either upload this data to an SAP support incident or use it to do your own troubleshooting.

13.2.4.2 View and Edit Extraction Results

Context

> → Remember
>
> Document Information Extraction typically provides extraction results for an average document in about 30 seconds.

However, processing can take longer if the task involved is more complex – for example, if the documents processed are large.

Before you use the service for important or time-sensitive tasks, we strongly recommend running mass tests to assess the performance of the service and make sure it meets your requirements.

Procedure

1. Click the *Document* icon in the left navigation pane.
2. Choose the row on the *Documents* screen with the document that you want to view. You now see the page preview of the document file.

   > ① Note
   >
   > If your device has a small screen, and you have difficulty checking the fields in the page preview, download the PDF document for full-screen display.

3. Click *Extraction Results* to see the results for header fields and line items.

   You can also see the machine learning model *Extraction Confidence Range* classified by colors: red (confidence between 0% and 50%), orange (confidence between 51% and 79%), and green (confidence between 80% and 100%).

   > ① Note
   >
   > You can set the lower threshold and change the color for each extraction confidence range under *UI Settings*. For further information, see Change Confidence Range Settings [page 248].

   To view the prediction confidence score for each header field and line item extracted, as well as the field name and description, hover over a field name, for example *Invoice Number*.

   Hovering over a field name also displays the raw value for that field – in other words, the value before postprocessing. Raw values can differ from extraction results. For example, if the *Delivery Date* field of a purchase order contains "ASAP", Document Information Extraction can't convert this text into a date and therefore returns a null value. Viewing raw values enables you to identify the content of fields that couldn't be extracted.

   > → Tip
   >
   > If the `label` property is defined for schema fields, user-friendly names for header fields and line items are displayed in the extraction results. For further information, see Add Fields to Schema Version [page 208].

   > ① Note
   >
   > If you've activated the *Schema Field Categories* and assigned fields to categories in the schema used for the extraction, the fields appear under their respective categories.
   >
   > If you've used the Enrichment Data API to create enrichment data entities, and if these entities match fields extracted from your document, the *Extraction Results* panel includes an *Enrichment Fields*

section. This section shows the enrichment data matched to extracted fields. If you don't use the Enrichment Data API, the *Enrichment Fields* section doesn't appear on the UI.

Hovering your cursor over the label of an enrichment field displays a tooltip showing the ID and confidence as well as the enrichment method used. Information about the method helps you quickly troubleshoot any issues with your master data. To find out more about the methods available, see Get Result [page 146].

4. If corrections are required, and the document status is "READY", you can edit the extraction results under *Header Fields* and *Line Items*.

   To download the unedited results, click ⬇ *(download icon)* and choose *csv*, *json*, or *txt*.

5. Click *Edit*.

> → Tip
>
> To avoid losing your work if there's an outage, activate *Autosave*. The service then saves your edits automatically every 10 seconds.

You can edit extracted values manually on the right of the screen. You can also select them from the page preview in the middle of the screen. To do the latter, hover your mouse over the page preview. The mouse pointer changes to a crosshair cursor. Position the cursor at the corner of the value that you wish to select. Then, hold down the left mouse button. Move the cursor diagonally to the opposite corner to draw a bounding box around the value you want to select. Select the appropriate header or line item field from the *Field* dropdown in the *Assign Field* dialog. Add or change the value, as necessary. If you choose a line item, set the number in the *Row Index* field. Make sure the number that you enter here matches the appropriate line item in the *Label* column on the right of the screen. Click *Apply* in the *Assign Field* dialog to confirm your edits.

> ⓘ Note
>
> To prevent Document Information Extraction from extracting unwanted or irrelevant characters, you can also draw bounding boxes around parts of the field values. In this case, you must edit the value so that it includes only the values in the bounding box. If you associate documents edited in this way with templates, the templates extract only those characters in the part of the field defined by the bounding box. This approach can be useful if you want to exclude punctuation from the extraction, for example.

> → Tip
>
> If you've uploaded your documents without a template, you can create a template here using the extraction values you've edited.
>
> For instructions on how to do so, see Create Template from Document Extraction Results [page 270].
>
> Note that this option is no longer available after you confirm the document.
>
> Alternatively, you can associate the document with an existing template by choosing *Add to Template*.

> → Remember
>
> If you associate a document with a template and then use that template to extract information from the same document, the extraction values can differ from the ones you entered and confirmed during editing.

> The technical reason for differences of this kind is that the Document Information Extraction UI extracts data based on heuristics and not on exact matching of bounding boxes.

6. Delete any bounding boxes that you don't need. In *Edit* mode, hover over the tooltip for the relevant bounding box in the page preview. Double-click the tooltip to display the *Assign Field* dialog and then choose *Delete* to remove the bounding box and its coordinates.

7. Save your changes.

   To download your edited results, click ⬇ *(download icon)* and choose *csv*, *json*, or *txt*.

8. You can also confirm the document here. To do so, choose *Edit* again and then choose *Confirm*. When you confirm documents, the prediction confidence score of all header and line item fields is set to 1.0 (100%).

> ⚠ Caution
>
> Do not confirm unreviewed documents, which can include incorrect extraction results. Once the document status changes from "READY" to "CONFIRMED", you can no longer change the extraction results.
>
> For additional considerations when you confirm documents, see Confirm Documents [page 255].

13.2.4.2.1 Confirm Documents

There are a few points to bear in mind when you confirm documents.

- SAP reserves the right to use confirmed documents in the reporting of accuracy values and for analytics.
- By default, Document Information Extraction doesn't use your documents to retrain the service's machine learning models. To allow SAP to use your documents for this purpose, set the `dataFeedbackCollection` configuration key at API level to *true*. A checkbox appears on the UI requesting your consent each time you confirm documents.
- If you allow SAP to use your documents for retraining, Document Information Extraction automatically checks them for any personally identifiable information (PII). If a document contains PII data, it isn't used for retraining. You can deactivate these checks by setting the `performPIICheck` subconfiguration at API level to *false*.

For further details of API-level settings, see Create Configuration [page 122] and Configuration Keys [page 124].

13.2.4.3 Delete Documents

Procedure

1. In the left navigation pane, click the *Document* icon.

2. On the *Documents* overview screen, select the documents you want to delete by choosing the relevant checkboxes.

    To select all the documents in the list, choose the checkbox above the table.

3. Click *Delete* and then click *OK* to delete the documents you selected. These documents are then removed from the *Documents* list.

    You can also delete individual documents by choosing *Delete* on the document detail screen.

> → Remember
>
> You can't delete documents that are associated with templates. In such cases, you must first navigate to the *Template* overview screen and dissociate the document from the template. For further details, see Add Documents and Activate/Deactivate Template [page 267].

13.2.5 Schema Configuration

Use this Document Information Extraction UI feature to create schemas containing data fields found in standard or custom document types. As an administrator, you can use these schemas as a basis for creating templates. End users must select a schema and can also select a corresponding template when adding documents.

Context

> ⓘ Note
>
> This feature is available only to users with the administrator role (role collection `Document_Information_Extraction_UI_Templates_Admin`).
>
> For additional information on using schemas, see the best practices under Schema Configuration: Best Practices [page 273].

A schema contains a list of header fields and line item fields representing the target information you want to extract from a particular type of document. You must select a schema when you add documents to the Document Information Extraction UI.

> → Tip
>
> The Document Information Extraction UI provides preconfigured SAP schemas for the following standard document types: purchase order, payment advice, and invoice. You can use these schemas unchanged to upload documents.
>
> You can't edit original SAP schemas. Always create a copy and then change the default fields, as required.

> ⓘ Note
>
> To extract text from images captured by camera, create a schema for a custom document type and use the OCR engine type *Scene Text*.

Extraction results for scene text appear in the API, not on the Document Information Extraction UI.

For details of extracted header fields and line items, see the following sections of the Document Information Extraction documentation:

- Extracted Header Fields [page 291]
- Extracted Line Items [page 300]

For information about limitations on extraction from tables, see Technical Constraints [page 288].

Use this feature to do the following:

- Create Schema [page 257]
- Edit Schema [page 258]
- Create Copy of Schema [page 259]
- Add Data Fields [page 259]
- Activate/Deactivate Schema [page 264]
- Delete Schema [page 265]

13.2.5.1 Create Schema

Procedure

1. Open the Document Information Extraction UI, as described in Subscribing to the Document Information Extraction UI [page 243].
2. In the left navigation pane, choose *Schema Configuration*.
3. In the top right of the screen, click *Create*.
4. Enter a name and optionally a description for the new schema.

   > → Tip
   >
   > To get better results for extractions with generative AI (available with the service plan **Document Information Extraction, premium edition** only), you can use the schema description value. This value is used in the prompt for document processing with large language models (LLMs).

5. Select the appropriate type of document.

   If you select *Custom* here, you must also select an OCR engine type. To extract text from images, select *Scene Text*; otherwise, select *Document*.

   > → Remember
   >
   > Extraction results for scene text recognition appear in the API, not on the Document Information Extraction UI.

6. Choose *Create*.

7. Choose the row containing your new schema to display the details pane. Here, you can add data fields and also edit, copy, activate/deactivate, or delete the schema, as described in the following sections.

> ⚠ **Restriction**
>
> You can't add data fields to schemas created with document type *Custom* and OCR engine type *Scene Text*.
>
> In schemas created using document type *Custom* and OCR engine type *Document*, you can add data fields. In this case, no default extractors are available.

13.2.5.2 Edit Schema

Procedure

1. In the left navigation pane, choose *Schema Configuration*.
2. On the *Configurations* screen, choose the row containing the schema you want to edit.

   You now see the schema details.
3. To change the schema, click the *Edit* button.
4. In the *Edit Schema* dialog, you can change the name of your schema and add, remove, or edit the description.

> ⚠ **Restriction**
>
> If a schema is currently active, deactivate it before editing. When you deactivate a schema, its status on the *Configurations* screen changes to "INACTIVE".
>
> You can't deactivate schemas that provide the basis for templates. Otherwise, any changes to the schema would affect the field definitions for the relevant templates.
>
> Before deactivating a schema of this kind, first deactivate all templates based on it and then delete them.
>
> Once you've completed your changes, activate the schema again.

13.2.5.3 Create Copy of Schema

Use this feature to copy SAP or custom schemas. SAP schemas support standard document types. You can use these preconfigured schemas unchanged to add documents and create templates. You can also copy and edit SAP schemas as a basis for configuring schemas of your own.

Procedure

1. In the left navigation pane, choose *Schema Configuration*.
2. Click 🗐 *(copy icon)* in the row of the schema you want to copy on the *Configurations* screen.

   In the *Copy Schema* dialog, the original schema name, followed by "_copy", appears automatically.
3. Edit the name as required and add an optional description.
4. Click the *Copy* button.

   The copy you've created now appears in the *Schemas* list, with the status "INACTIVE".

13.2.5.4 Add Data Fields

Find out how to add header fields and line items to your schemas.

Procedure

1. In the left navigation pane, choose *Schema Configuration*.
2. On the *Schemas* screen, choose the row containing the schema you want to add data fields to.
   You now see the schema details.
3. If the schema has the status "ACTIVE", you must deactivate it before you can add data fields. In this case, click *Deactivate*.

   > ⚠ Restriction
   >
   > You can't deactivate schemas that provide the basis for templates. Otherwise, any changes to the schema would affect the field definitions for the relevant templates.
   >
   > Before deactivating a schema of this kind, first deactivate all templates based on it and then delete them.

4. To add a header field to the schema, click the *Add* button for *Header Fields*.
5. On the *New Field* panel, enter the name of the header field you want to extract, an optional field label, and an optional description.

> **→ Tip**
>
> Use the *Label* option to define user-friendly names for header and line item fields. Any field labels that you enter here replace the technical field names under *Extraction Results* in the *Documents* feature of the Document Information Extraction UI.

> **ⓘ Note**
>
> If you've enabled the *Schema Field Category* feature, you can also assign the data field to an appropriate category. For more information, see Assign Schema Fields to Categories [page 264].

6. Next, select the data type – either *country/region*, *currency*, *discount*, *date*, *number*, or *string*.

> **→ Remember**
>
> The data type *country/region* extracts the values in a two-letter code (alpha-2) ISO 3166 format. For example, DE for Germany, FR for France, GB for United Kingdom, and US for United States.

7. In the *Setup Type* dropdown, use the prefilled value (*auto* or *manual*) or change it in line with your needs.

> **ⓘ Note**
>
> Which setup type you select here depends on a number of factors, including document type, preferred extraction method, and which service plan you're using.
>
> For details of setup types and associated factors, see Setup Types [page 261].

8. Click *Add*.
   On the *Configurations* panel on the left of the screen, the status of the schema changes to "DRAFT".

9. If you want to edit the data field, click either the link in the *Field Name* column or the ✏ *(edit icon)* in the relevant row under *Action*.

10. To add line item fields to the schema, click the *Add* button for *Line Item Fields*.

11. Enter the data for the new line item field in the same way as you did for the header field.

Related Information

Setup Types [page 261]

13.2.5.4.1 Setup Types

Learn about the setup types available when you add data fields to schemas. Find out how these setup types relate to document types, extraction methods, and default extractors.

Available Setup Types

When you add data fields to a schema on the Document Information Extraction UI, you can select one of the following setup types:

- *auto*
- *manual*

These setup types support extraction using different methods, depending on whether the schema was created for a standard or for a custom document type.

Default Values

When you first call up the *Add Data Fields* dialog, the service prefills the *Setup Type* field. The default values depend on the document type and which edition of Document Information Extraction you use:

- Premium edition
  - Schemas for standard and custom document types: *auto*
- Base edition
  - Schemas for standard document types: *auto*
  - Schemas for custom document types: *manual*

You can change these prefilled values in line with your needs.

Document Types, Setup Types, Extractions Methods, and Default Extractors

The following table shows the various combinations of document type and setup type and how they relate to the extraction method and the use of default extractors:

Document Type for Schema	Setup Type	Extraction Method	Select Default Extractor?
Standard	*auto*	Service's machine learning models	Yes
		Generative AI	Not applicable

Document Type for Schema	Setup Type	Extraction Method	Select Default Extractor?
	manual	Template	Not applicable
Custom	*auto*	Generative AI	Not applicable
	manual	Template	Not applicable

> ⚠ **Restriction**
>
> The setup type *auto* without default extractor (extraction method: generative AI) is available only with the service plan **Document Information Extraction, premium edition** (`premium_edition`). See Service Plans [page 84] and Metering and Pricing [page 86].
>
> However, if you want to try out extraction using generative AI, you can do so with an SAP BTP trial account. Simply follow the steps in the tutorial: Use Trial to Extract Information from Custom Documents with Generative AI and Document Information Extraction⚓

> → **Remember**
>
> You can use different extraction types for header fields in the same schema. However, you can't combine different extraction types for line items in the same schema.
>
> For example, if you use the setup type *auto* without a default extractor for one line item field, you must use it for all the other line item fields that you add to your schema.

> ⚠ **Caution**
>
> Always validate information extracted using generative AI before using it for critical applications.
>
> If you prefer not to use generative AI to extract information from documents, select the setup type *auto* with a default extractor (standard document types only). Alternatively, select the setup type *manual* (standard and custom document types) when adding data fields to your schema.

> ⓘ **Note**
>
> As of October 9, 2023, the setup type *default* is no longer available for new schemas. If an existing schema includes fields added before this date with the setup type *default*, you can use only this setup type when adding new fields. Schemas created before this date that don't yet include any fields offer you the choice of *auto* or *manual* as setup type.
>
> Because SAP schemas include fields added before October 9, 2023, when you copy these schemas, the only setup type available is *default*.

Related Information

Extraction Using Generative AI: Languages [page 101]
Add Fields to Schema Version [page 208]
Extraction Using Generative AI: Best Practices [page 286]

13.2.5.5 Schema Field Categories

Find out about schema field categories, how they help you manage data fields more easily, and how to enable the relevant feature in the Document Information Extraction UI.

If your schemas include many data fields, you can use the *Schema Field Categories* feature to group data fields. Using categories means that you no longer have to scroll through a long list to find the specific fields you're looking for. Instead, you can quickly and easily pinpoint and edit fields by finding and expanding the relevant category.

> ⓘ Note
>
> The *Schema Field Categories* feature isn't enabled by default. If you wish to use it, you must first enable it. For guidance on how to do so, see Enable Features [page 247].

13.2.5.5.1 Create Schema Field Categories

Create a category to group together schema data fields.

Prerequisites

You've enabled the *Schema Field Categories* feature, as described in Enable Features [page 247].

Context

If your schemas include many different data fields, you can create categories to group data fields. Doing so avoids long lists of field names and makes it easier to find specific fields.

Procedure

1. On the left navigation pane, choose *Schema Configuration*.
2. On the *Configurations* screen, choose the row containing the schema you want to add field categories to. You now see the schema details.
3. If the schema has the status "ACTIVE", you must deactivate it before you can add categories. In this case, click *Deactivate*.

   > ⚠ Restriction
   >
   > You can't deactivate schemas that provide the basis for templates. Otherwise, any changes to the schema would affect the field definitions for the relevant templates.

> Before deactivating a schema of this kind, first deactivate all templates based on it and then delete them.

4. To add a category to the schema, click the *Add* button for *Category*.

5. On the *New Category* panel, enter a suitable name for your new category, an optional description, and an optional label.

6. Save your entries.

   You can see the name of your new category as a link in the *Category Name* column. Clicking on this link allows you to view the category details, edit them, or delete the category.

   You can also use the icons in the *Action* column to edit or delete categories.

13.2.5.5.2 Assign Schema Fields to Categories

Find out how to group data fields in schemas by category.

You've enabled the *Schema Field Categories* feature, as described in Enable Features [page 247], and have created at least one category.

With *Schema Field Categories* enabled, the steps for creating, viewing, and editing data fields are largely the same as the ones described in Add Data Fields [page 259].

The main difference is that when you create, view, or edit a header field or line item, the panel on the right of the screen includes the *Category* dropdown. You can use this dropdown to select the category that you wish to assign your data field to.

If your schema includes many data fields, assigning fields to categories eliminates long lists, making it easier to find individual fields.

13.2.5.6 Activate/Deactivate Schema

Procedure

1. In the left navigation pane, choose *Schema Configuration*.

2. On the *Configurations* screen, choose the row containing the schema you want to activate.

   You now see the schema details.

3. To activate the schema, click the *Activate* button. On the *Configurations* screen, the schema status changes to "ACTIVE".

   If a schema doesn't yet have any data fields, the *Activate* button is grayed out.

4. When a schema has the status "ACTIVE", the *Deactivate* button replaces the *Activate* button.

> ⓘ Note
>
> If you wish to change or delete a schema that is active, you must first click *Deactivate*. When you deactivate a schema, its status on the *Configurations* screen changes to "INACTIVE". To enter your changes, choose *Edit* ✎ *(pen icon)* Once you've completed your changes, activate the schema again.

> ⚠ Restriction
>
> You can't deactivate schemas that provide the basis for templates. Otherwise, any changes to the schema would affect the field definitions for the relevant templates.
>
> Before deactivating a schema of this kind, first deactivate all templates based on it and then delete them.

13.2.5.7 Delete Schema

Procedure

1. In the left navigation pane, choose *Schema Configuration*.
2. On the *Configurations* screen, select the checkbox for the row containing the schema you want to delete.

   You can't delete a schema that has the value "YES" in the *SAP Schema* column.
3. If the schema has the status "ACTIVE", you must deactivate it before you can delete it. In this case, click *Deactivate*.

   > ⚠ Restriction
   >
   > You can't deactivate schemas that provide the basis for templates. Otherwise, any changes to the schema would affect the field definitions for the relevant templates.
   >
   > Before deactivating a schema of this kind, first deactivate all templates based on it and then delete them.

4. Click *Delete* and then *Yes* to delete the selected schema.

   The schema is removed from the *Schemas* list.

13.2.6 Template

Use this Document Information Extraction UI feature to create, reuse, edit, and delete templates based on schemas and document types. End users can select templates together with a corresponding schema to extract information from business documents of the appropriate type and structure.

Context

> ⓘ Note
>
> This feature is available only to users with the following administrator role:
>
> - `Document_Information_Extraction_UI_Templates_Admin`
>
> For additional information on using templates, see the best practices under Template [page 278].

Templates are based on schemas and enable you to show the position of extraction fields in a particular document layout. After creating a template, you use the *Document* feature to associate one or more documents with it. You then edit the extraction results for these documents, indicating the location of fields and their values.

Templates are essential for processing custom document types. However, you can also use them with standard document types to fine-tune extraction results.

> → Tip
>
> If you follow the guidance in General Recommendations and Limitations [page 279], you only have to edit the extraction results for one document that you associate with your template.

Use this feature to do the following:

- Add Template [page 266]
- Add Documents and Activate/Deactivate Template [page 267]
- Export/Import Template [page 270]
- Create Template from Document Extraction Results [page 270]
- Delete Template [page 271]

13.2.6.1 Add Template

Procedure

1. Open the Document Information Extraction UI, as described in Subscribing to the Document Information Extraction UI [page 243].

2. Click the *Template* icon in the left navigation pane.

3. Click *Create a new template* (✛ *add icon*) at the top right.

4. Enter a name and optionally a description for the new template. Select the appropriate document type (either *Invoice*, *Payment Advice*, *Purchase Order*, or *Custom*). Choose the schema you wish to use as a basis for the new template. Click *Create*.

5. Choose *OK* to see the template details.

   The *Extraction Fields* tab shows the header fields and line item fields from the schema you specified.

6.   ⓘ Note

   This step and the ones that follow are optional and are only applicable if you want to assign a fixed value to one or more extraction fields.

   Choose the *Extraction Fields* tab and then choose *Edit* on that tab.

7. Enter a value that you wish to associate with all instances of a particular field.

   For example, if you intend to use your template only for documents from one supplier, you could enter the name of that supplier as the fixed value for the *senderName* field.

8. Repeat the preceding step for any other fields that you want to assign fixed values to.

9. Save your entries.

13.2.6.2 Add Documents and Activate/Deactivate Template

Context

To add documents to a template, you use the *Document* feature of the Document Information Extraction UI. Adding documents to templates, as described here, helps improve accuracy.

⚠ Restriction

The document and the template that you wish to add it to must share the same schema. If the document and template have different schemas, you can't add the document to the template.

Procedure

1. Choose the *Document* icon in the left navigation pane.

2. Add a document or documents as described in the chapter Add Document [page 251].

3. Choose the row with the document that you want to work with.

   You now see the document details. It's best if the file has at least 2 line items.

4. Edit the extraction results for the document as described in View and Edit Extraction Results [page 252].

    You can confirm the document at this point. It's not necessary to save the document. When you associate a document with a template, the Document Information Extraction UI saves the extraction results automatically.

    > → Remember
    >
    > If you associate a document with a template and then use that template to extract information from the same document, the extraction values can differ from the ones you entered and confirmed during editing.
    >
    > The technical reason for differences of this kind is that the Document Information Extraction UI extracts data based on heuristics and not on exact matching of bounding boxes.

5. To add this document to a template, choose *Add to Template* at the top of the pane on the right of the screen.
6. Select the relevant template from the dropdown and choose *Add*.

    The document file is added to the template that you selected. It's displayed as an associated document on the details page for this template.
7. Repeat the preceding steps to add more documents to your template.

    > ⚠ Restriction
    >
    > You can add a maximum of 5 documents to a template.

8. If you want to remove associated documents from a template, first choose the *Template* icon in the left navigation pane.
9. Then select the relevant template.
10. Choose the 🔗 *(broken link)* icon in the *Action* column of the *Associated Documents* tab.
11. Finally, choose *OK* to confirm the action.
12. Activate a template in status "DRAFT" to use it to extract results from documents similar to the ones associated with it.

    The template status changes from "DRAFT" to "ACTIVE".
13. Deactivate a template in status "ACTIVE" to edit it, delete it, or make it no longer available for Document Information Extraction.

    The template status changes from "ACTIVE" to "DRAFT".

13.2.6.3 Edit Template

Find out how to make changes to templates.

Context

If you want to make changes to a template, you can do so using the *Edit* function. You can change the template name and description. In addition, you can select a different schema for the template. Changing the schema makes a new set of extraction fields available for the template.

> ⚠ **Restriction**
>
> If a template is currently active, you must deactivate it before you can edit it.

Procedure

1. Click the *Template* icon in the left navigation pane.
2. Select the template that you want to edit.
3. Click *Edit*.

   The *Edit Template* dialog appears. Here, you can change the name and description by editing the corresponding fields.

   You can also select a different schema for your template. To change the schema, do the following.

4. Choose the *Schema* dropdown and select a schema from the list.

   > ⓘ **Note**
   >
   > This list includes only schemas that match the document type for which the template was originally created.

5. Click *Save* to complete your changes.

   > → **Remember**
   >
   > If you've already edited extraction result for sample documents associated with your template, these edits are preserved following the change of schema only for fields that appear in both the old and the new schema. After changing the schema, you can annotate the newly added fields in your existing sample documents.

13.2.6.4 Export/Import Template

Avoid duplicated effort by reusing existing templates in different clients.

Context

You've created a template in a test client by following the steps in Add Template [page 266] and Add Documents and Activate/Deactivate Template [page 267]. You're now happy with your new template and want to export it from the current client before importing it into your production client.

The steps presented here assume the following:

- You're still in the test client.
- You've selected your new template from the *Templates* list and are now on the screen showing the template details.

Procedure

1. Choose *Export*.

   Document Information Extraction downloads the template to your local machine. The download includes the `schema.json` and `template.json` files and a folder with the associated documents.

2. Choose *Change Client* and select the production client to which you want to import your template.

   The Document Information Extraction UI displays the *Templates* list for the production client.

3. Choose ⬆ *(upload icon)* and navigate to the folder you downloaded in Step 1.

4. Select the folder and choose *Open*.

   The new template appears in the list. Users can now select this template when adding documents of the appropriate type to the Document Information Extraction UI .

13.2.6.5 Create Template from Document Extraction Results

This feature allows you to quickly and easily create templates when adding documents to the Document Information Extraction UI.

Context

You've added a document by following the steps in Add Document [page 251] and View and Edit Extraction Results [page 252].

> → Remember
>
> To create a template based on document extraction results, you must use a schema when adding the document.

Before creating a template from the document extraction results, make sure that the document you want to base your template on has the status "READY".

Procedure

1. Choose *Create Template* on the *Documents* UI.
2. Enter a name for your template (mandatory) and a description (optional), then choose *Create*.

   The template detail screen appears, showing your new template with the preprocessing status "DONE".

   You can now use your template in the same way you'd use one created directly using the *Template* feature.
3. Activate, edit, export, or delete your template, as described in Add Documents and Activate/Deactivate Template [page 267], Export/Import Template [page 270], and Delete Template [page 271].

13.2.6.6 Delete Template

Procedure

1. In the left navigation pane, click the *Template* icon.
2. On the *Templates* screen, click the row containing the template you want to delete.

   You see the template details.

   > ⚠ Restriction
   >
   > If a template is currently active, you must deactivate it before you can delete it.
3. Click *Delete* and then *OK* to delete the selected template.

   The template is removed from the *Templates* list.

14 Best Practices

Find out about recommended approaches for optical character recognition, the main features of the Document Information Extraction service, data enrichment, and extraction using generative AI.

The quality of your extraction results depends on a wide range of factors. This section is intended to help you get the best out of the Document Information Extraction service. It includes the following information:

- General recommendations on how to get better extraction and enrichment results using OCR best practices.
- Decision procedures, recommendations, and tips on how to use the schema configuration, template, and document features of Document Information Extraction.
- Important considerations when using generative AI to extract information from documents automatically.

Related Information

14.1 Optical Character Recognition (OCR): Best Practices

To get better extraction and enrichment results, bear in mind the following when uploading document files to the Document Information Extraction service:

- Use page size A4 (Europe) or letter (United States).
- Portrait orientation is preferable.
- Use a high-quality scan.
- A handwriting detection feature is available. At present, this feature detects only handwriting in English.
- The ideal resolution is 300 dpi. For good quality, at least 150 dpi is needed. Higher resolution (> 300 dpi) generally has no effect on extraction results. Be aware that very large files take longer to preprocess because they're scaled back to 300 dpi. In addition, the service ignores colors and converts images to grayscale.
- Make sure that the text isn't blurred.
- The service extracts dark text on a light background more accurately than light text on a dark background.
- Avoid handwritten additions, such as texts, numbers, checkmarks, or underlining, as well as highlighting with marker pens. Additions of this kind can lead to poor OCR and extraction results.

- Words that are oriented differently (for example, rotated 90 degrees) or have a much larger or much smaller font than those on the rest of the page aren't detected.
- Very small fonts can result in poor OCR and extraction results.
- Where text takes up only a small area of the page, excessive zoom-in or cropping can cause extraction issues.

> → Tip
>
> - See Supported Document Types and File Formats [page 91].
> - See also Supported Languages and Countries/Regions [page 93] and Technical Constraints [page 288].

14.2 Schema Configuration: Best Practices

Learn about best practices for using schemas to upload documents to the Document Information Extraction UI.

You must always use a schema when uploading documents to the Document Information Extraction UI.

> ⓘ Note
>
> To use the *Schema Configuration* feature to create, copy, and edit schemas, you must have the administrator rights provided by the following role collection:
>
> - `Document_Information_Extraction_UI_Templates_Admin`
>
> If you have the `Document_Information_Extraction_UI_End_User` role, you can use any available schemas, except SAP schemas, to upload documents.

The steps involved in adding a schema differ depending on whether the document type is standard or custom. For details of the respective processes, see the subtopics in this section.

Related Information

Standard Document Types [page 273]
Custom Document Types [page 276]

14.2.1 Standard Document Types

Configure schemas for standard document types.

The Document Information Extraction UI supports the following standard document types:

- Invoice

- Payment advice
- Purchase order

The following image outlines the steps and settings for processing standard document types with or without a template.
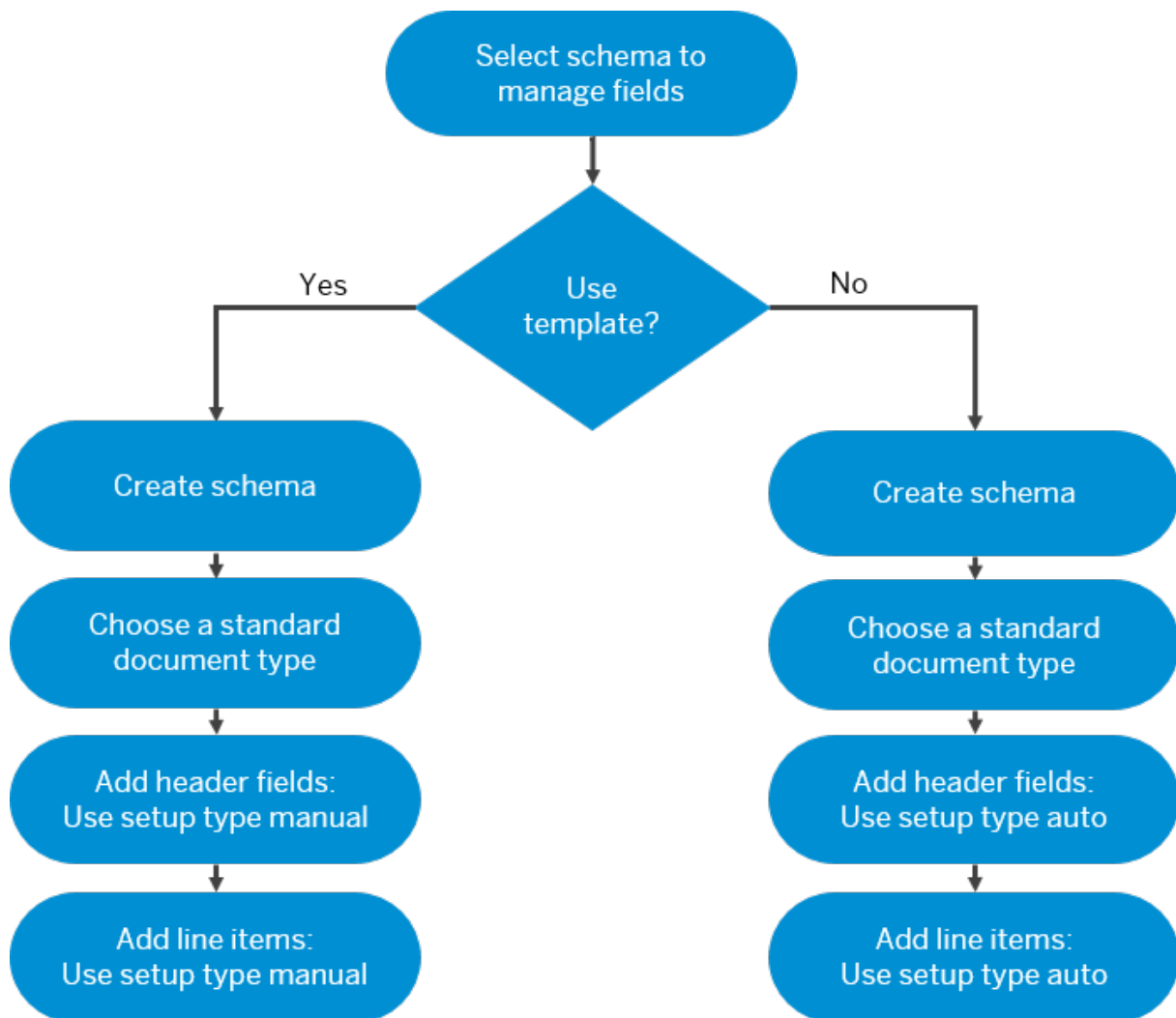
- Template: Best Practices [page 278]
- Create Schema [page 257]
- Create Schema [page 257]
- Setup Types [page 261]
- Setup Types [page 261]
- Add Document [page 251]

> → Remember
>
> SAP schemas provide a set of typical fields with default extractors for standard document types. If you don't want to configure schemas for standard document types from scratch, you can select the appropriate SAP schema unedited when you add a document or create a template on the Document Information Extraction UI. No configuration is needed when you use SAP schemas in this way.
>
> You can also create your own schema by copying the SAP schema for the relevant standard document type. You can then edit this copy, choosing some or all the fields from the SAP schema as a basis for your own schema and adding custom fields, as required.

Extraction Methods, Setup Types, and Default Extractors

You can use the following extraction methods for header fields in schemas for standard document types:

- Template: setup type *manual* without default extractor
- Machine learning models of the Document Information Extraction service: setup type *auto* with default extractor
- Generative AI: setup type *auto* without default extractor

> ⚠ Restriction
>
> The generative AI extraction method is available only with the service plan **Document Information Extraction, premium edition** (`premium_edition`).

> → Remember
>
> You can use different extraction types for header fields in the same schema. However, you can't combine different extraction types for line items in the same schema.
>
> For example, if you use the setup type *auto* without a default extractor for one line item field, you must use it for all the other line item fields that you add to your schema.

Default Extractors

Templates generally deliver better results for custom header fields than for custom line items. To get the best extraction results when using a template or the machine learning models of the Document Information Extraction service with standard document types, configure default extractors for header and line item fields as follows:

- Header fields: Don't use default extractors for custom header fields. You can then use a template to edit them.
- Line items: Use default extractors, wherever possible.

To access the default extractors when configuring a schema, choose *Header Fields* or *Line Item Fields* and then choose *Add*. Next, select the relevant data type and the setup type *auto*. You can now select the appropriate extractor for the data type from the *Default Extractor* dropdown.

Related Information

14.2.2 Custom Document Types

Configure schemas for custom document types.

Custom documents are documents that don't belong to the standard document types in Document Information Extraction. There are many different types of custom document: Common examples include powers of attorney, birth certificates, and résumés.

The following image outlines the steps and settings for processing custom document types with and without a template.
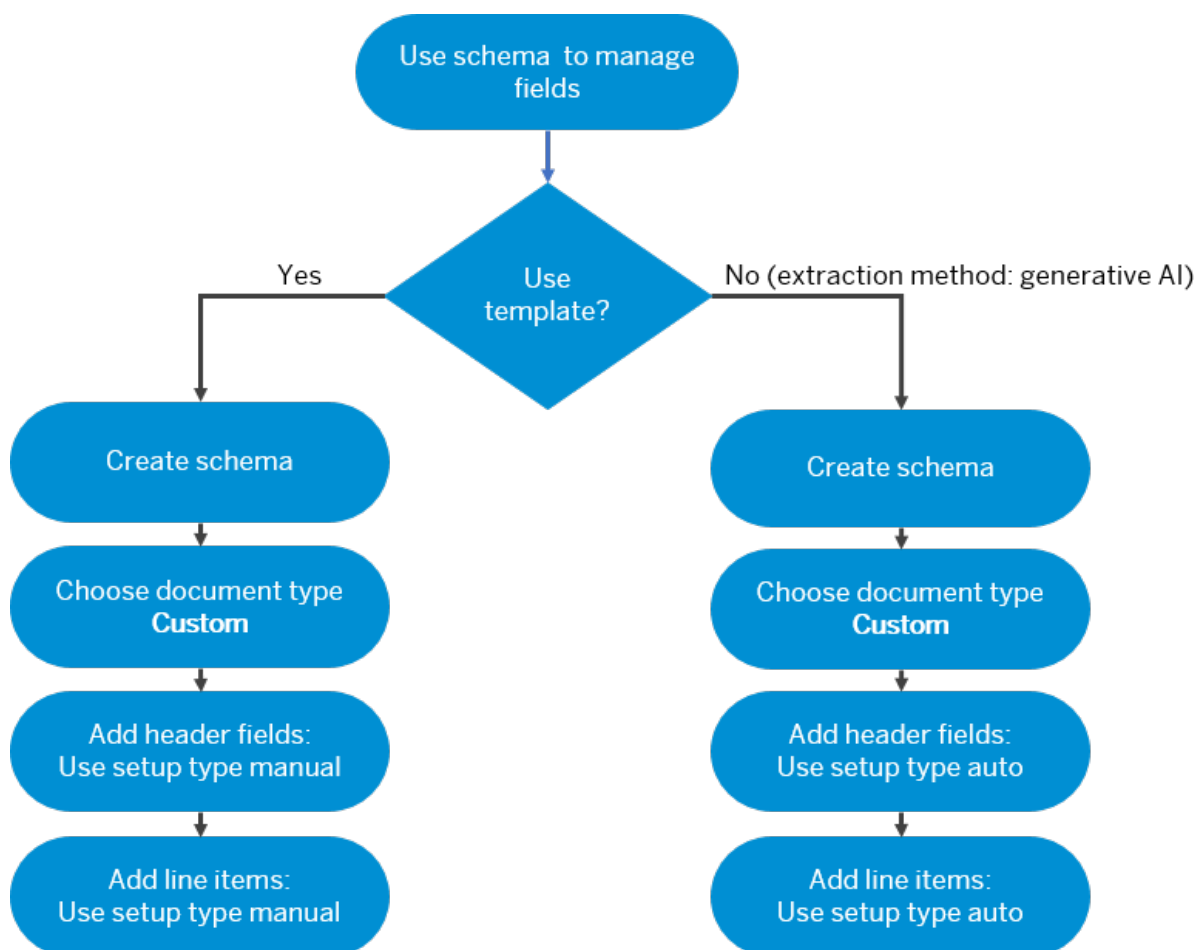
Extraction Methods and Setup Types

You can use the following combinations of extraction methods and setup types for header fields in schemas for custom document types:

- Template: setup type *manual*
- Generative AI: setup type *auto*

> ⚠ **Restriction**
>
> The generative AI extraction method is available only with the service plan **Document Information Extraction, premium edition** (`premium_edition`).

> ⓘ **Note**
>
> Default extractors aren't available for custom document types.

> → **Remember**
>
> You can use different extraction types for header fields in the same schema. However, you can't combine different extraction types for line items in the same schema.
>
> For example, if you use the setup type *auto* for one line item field, you must use it for all the other line item fields that you add to your schema.

Related Information

Add Data Fields [page 259]
Setup Types [page 261]

14.3 Template: Best Practices

Decide whether to use a template when uploading documents to the Document Information Extraction UI and make the relevant settings.

When uploading documents to the Document Information Extraction UI, you must select a schema. In addition, you have the option of using a template. The associated decision procedure is as follows:
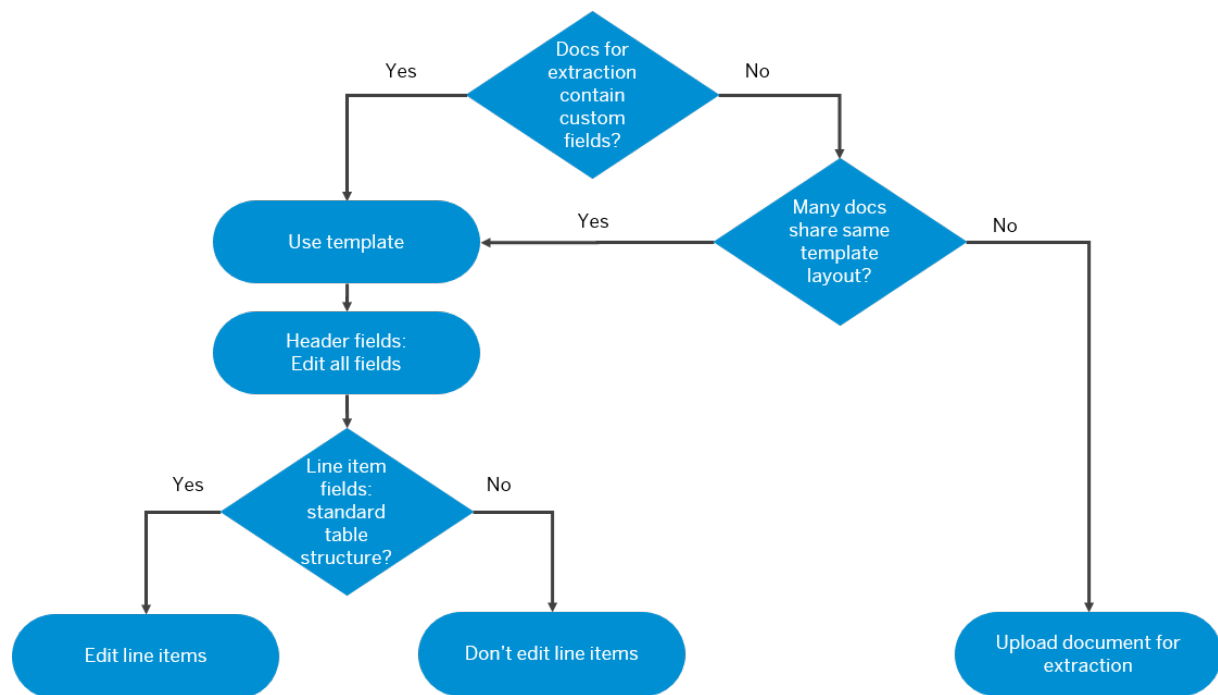
- Template: Best Practices [page 278]
- Standard and Custom Tables [page 280]
- Add Document [page 251]

> ⓘ Note
>
> To use the *Template* feature to create templates, you must have the administrator rights provided by the following role collection:
>
> - `Document_Information_Extraction_UI_Templates_Admin`

The Document Information Extraction UI delivers best results with standard table structures. If your documents include custom fields, we recommend using a template. This approach allows you to edit extraction results for fields that don't have default extractors. Edit all custom header fields. If the line items in your documents are in a standard table structure, also edit the line items. However, if the table has a custom structure, don't edit the line items.

If the documents don't include custom fields, and only a few of the documents share the same template layout, don't use a template. In this case, upload the documents using a schema only.

If your documents don't include custom fields, but many of them share the same template layout, use a template. If the line items in your documents are in a standard table structure, edit the line items. However, if the table has a custom structure that is likely to cause issues with the template approach, don't edit the line items.

> ⓘ **Note**
>
> If there are extraction errors when using templates, refer to the subsections of these template best practices.

Related Information

General Recommendations and Limitations [page 279]
Standard and Custom Tables [page 280]

14.3.1 General Recommendations and Limitations

Follow best practices and be aware of limitations when using templates to extract information from custom and standard document types.

Templates are essential when extracting information from custom document types, for which Document Information Extraction has no pre-trained models. In addition, templates can help you fine-tune results when extracting information from standard document types. (See Standard Document Types [page 273].)

Whether you use templates to extract information from custom or standard document types, note the recommendations here and in Standard and Custom Tables [page 280]:

- Use templates only with well-structured form-like documents such as the following: structured forms, application forms, certificates, prescriptions, and personal IDs.
- If possible, process one-page documents only. Otherwise, the results can be less accurate.

Note the following constraints on header and line item fields:

- If the same header field appears on more than one page, the Document Information Extraction UI extracts this field only once.
- Templates support multiple tables per page, provided they all have a standard structure and the same table headers. Multiple tables that are horizontally placed aren't supported.
- Nested table structures (with items grouped in the same line) cause issues.
- Items that overlap horizontally (for example, different items in the same column) also cause problems.
- Header and line item fields with identical or very similar formatting prevent the template from distinguishing the header from the main part of the table. As a result, the template can't detect where the table starts.
- If adjacent columns are too close to each other, the Document Information Extraction UI can't distinguish them. In such cases, the service extracts the contents of multiple columns as a single value.

> **⚠ Caution**
>
> If there are extraction errors when using templates, check for the following issues:
>
> - Document for upload has significant page rotation/tilt (15 degrees or more).
> - Size of pages and margins differs between document for upload and associated document.
> - Position of image differs between document for upload and associated document.
> - Line items in the document for upload differ slightly from the line items in the associated document.
> - Images include scanning noise – for example, background images and bleed through, where text on the back of the document is visible on the front.
> - OCR results are poor.
>
> These issues result in fields failing to map to their expected positions. In such cases, extraction can either be incorrect (wrong value) or fail entirely (no value). If extraction fails, the system falls back to the pre-trained global model, which can result in incorrect extraction.

Related Information

14.3.2 Standard and Custom Tables

Compare the tables in your documents with examples of standard and custom structures.

If you use a template to extract information from tables, you get the best results from simple, well-structured layouts (standard tables). By contrast, custom tables can cause issues.

Before using a template, compare the tables in your documents with the following examples of standard and custom tables.

> **→ Remember**
>
> Whether you're extracting information from standard or custom tables, bear the following layout-related points in mind:
>
> - If you use a template, make sure that the header and line item fields are formatted differently from each other. If they have very similar or identical formatting, the template can't distinguish the header from the main part of the table and therefore can't detect where the table starts.
> - Make sure that adjacent table columns aren't too close to each other. If they are, the Document Information Extraction UI can't distinguish them. As a result, it extracts the contents of multiple columns as a single value.

Standard Tables

For best results, use tables with the standard structures shown here.

In the following examples, the column headings correspond to the header fields, and the line items appear directly under them.

Headers Arranged Horizontally from Left to Right

Material number	Description	Quantity	Unit price	Total price
123	Product 1	1	EUR 12.35	EUR 12.35
234	Product 2	2	EUR 2.35	EUR 2.35

Headers Arranged Horizontally from Left to Right: No Nested Structures

Material number	Description	Quantity	Unit price	Total price
123	Product 1 Description covering several lines	1	EUR 12.35	EUR 12.35
234	Product 2 Description covering several lines	2	EUR 2.35	EUR 2.35

As shown in both of the preceding tables, headers are arranged horizontally from left to right in standard tables. If a column includes content that covers more than one line (as in the **Description** column of the second table), this content isn't nested. In other words, it's not spread across multiple columns.

See the contrasting examples in the **Custom Tables** section.

Custom Tables

Tables structured as shown in this section can cause issues during extraction and deliver poorer results.

Headers Arranged Vertically

Material number	123	234
Description	Product 1	Product 2
Quantity	1	2
Unit price	EUR 12.35	EUR 2.35
Total price	EUR 12.35	EUR 2.35

Nested Structures

```
--- I N V O I C E ---

                                              Date - Sat 3rd. Jul 2021

Week Ending - Sat  5th. Jun 2021          ------------ Owing from Previous Period =    29.80
                                                        Less Payments Made =

   Item        -      Sun  Mon  Tue  Wed  Thu   Fri  Sat    Qty        Price
   Semi Skimmed1Pt Glas  CAN  CAN  CAN  CAN  CAN  CAN    4      4 @    0.70  =      2.80
   6 Large Eggs          CAN  CAN  CAN  CAN  CAN  CAN    1      1 @    1.30  =      1.30
                                                              Weeks Total =       £4.10
                                                                Sub Total =        4.10

Week Ending - Sat 12th. Jun 2021          ------------ Owing from Previous Period =     4.10

   Item        -      Sun  Mon  Tue  Wed  Thu   Fri  Sat    Qty        Price
   Semi Skimmed1Pt Glas            4        3         4     11 @    0.70  =      7.70
   6 Large Eggs                                       1      1 @    1.30  =      1.30
                                                              Weeks Total =       £9.00
                                                                Sub Total =       13.10

Week Ending - Sat 19th. Jun 2021          ------------ Owing from Previous Period –    13.10

   Item        -      Sun  Mon  Tue  Wed  Thu   Fri  Sat    Qty        Price
   Semi Skimmed1Pt Glas            4        3         4     11 @    0.70  =      7.70
   6 Large Eggs                                       1      1 @    1.30  =      1.30
                                                              Weeks Total =       £9.00
                                                                Sub Total =       22.10

Week Ending - Sat 26th. Jun 2021          ------------ Owing from Previous Period =    22.10

   Item        -      Sun  Mon  Tue  Wed  Thu   Fri  Sat    Qty        Price
   Semi Skimmed1Pt Glas            4        3         4     11 @    0.70  =      7.70
   6 Large Eggs                                       1      1 @    1.30  =      1.30
                                                              Weeks Total =       £9.00
                                                                Sub Total =       31.10

Week Ending - Sat  3rd. Jul 2021          ------------ Owing from Previous Period =    31.10

   Item        -      Sun  Mon  Tue  Wed  Thu   Fri  Sat    Qty        Price
   Semi Skimmed1Pt Glas            4        3         4     11 @    0.70  =      7.70
   6 Large Eggs                                       1      1 @    1.30  =      1.30
                                                              Weeks Total =       £9.00
                                                                Sub Total =       40.10
```

Items Overlapping Horizontally

Artikel	Product/Eenheid	Korting Hoeveelheid	Eenheidsprijs	Totale Prijs	BTW	Uw Ref
000010	Z0251-500G	1,0	EUR	EUR		
- 10						
	ZINC SULFATE HEPTAHYDRATE CELL CULTURE*T	1	143,00 /EA		21,0	
% korting		5,0-%	7,15-	135,85		
	Estimated delivery date	04.02.2021				
000020	429430-5G	2,0	EUR	EUR		
- 20						
	Zinc chloride, anhydrous, powder, 99.995	2	83,00 /EA	166,00	21,0	
	Estimated delivery date	05.02.2021				

→ Tip

If your documents include custom tables, we recommend using default extractors for all line items when configuring the corresponding schema. If you then decide to use the *Template* function with your schema, you don't have to edit the extraction results for the line items.

ⓘ Note

If you follow the guidance in this subsection but still have extraction errors, refer to the general recommendations for using templates.

General Recommendations and Limitations [page 279]

14.4   Document: Best Practices

Make the recommended settings for uploading documents to the Document Information Extraction UI.

You must always use a schema when uploading documents to the Document Information Extraction UI. Schemas enable you to manage fields for extraction centrally, reducing manual effort, and inconsistencies.

When you add documents, the decision procedure is as follows:
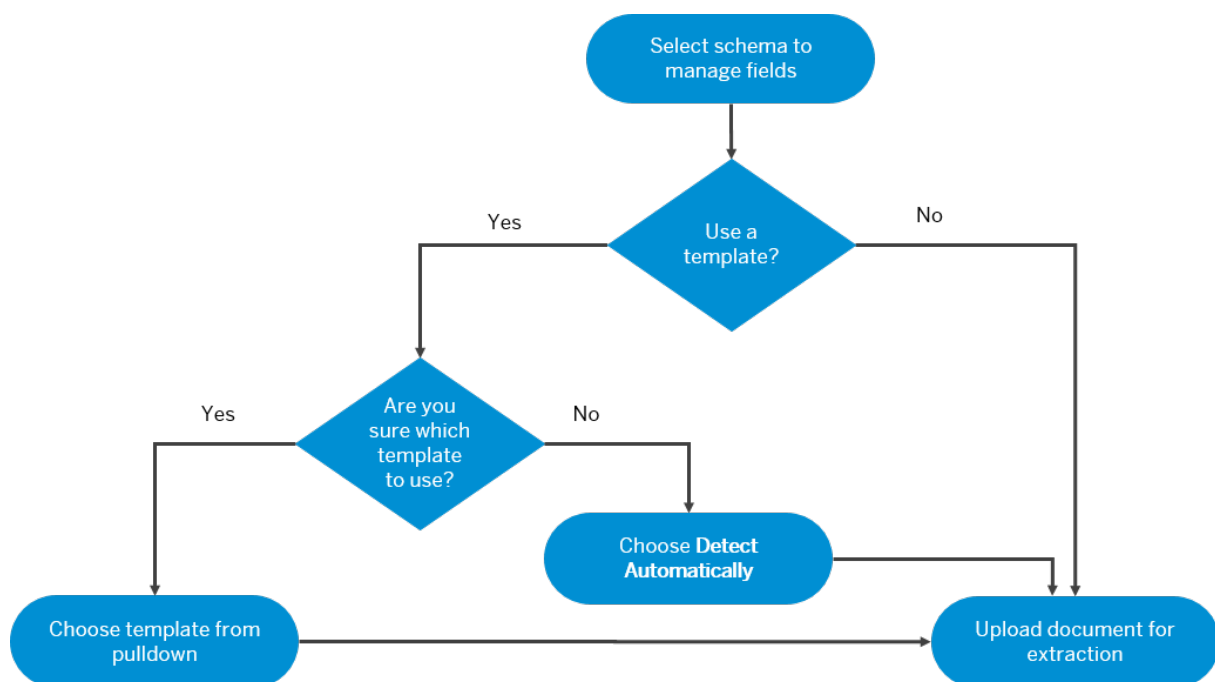
- Add Document [page 251]
- Schema Configuration: Best Practices [page 273]
- Template: Best Practices [page 278]

If you want to use a schema without a template, simply select the appropriate schema and then upload your documents to the Document Information Extraction UI.

If you want to use a schema with a template and know the template name, select the template from the dropdown in the *Select Document* step. If you're unsure which template to use, choose *Detect Automatically*. The service then finds the best template for your document.

> **→ Tip**
>
> When uploading documents using a schema, you may find that a suitable template isn't available. In this case, you can create a template based on the extraction results for your documents. For details of how to do so, see Create Template from Document Extraction Results [page 270].
>
> To create templates in this way, you need the admin rights provided by the following role collection:
>
> - `Document_Information_Extraction_UI_Templates_Admin`

14.5 Data Enrichment: Best Practices

Data enrichment is a powerful feature that matches vendors, customers, employees, and products found on a document with master data uploaded to the Document Information Extraction service.

To improve the performance of the data enrichment feature, make sure that your master data is up to date and activated. To get the best possible matching results, observe the following recommendations:

- Don't use placeholder values for individual fields that lack a value. Remove these fields instead.
- Always include the keys `name` and `address1` and populate them with a valid supplier or customer name and address. Otherwise, the enrichment is unlikely to work as intended.
- Whenever possible, include `taxId` and `bankAccount` information in the `businessEntity` field. These two fields have benefits for the enrichment.
- Always keep in mind that uploaded master data must be activated before it can be used for enrichment. If automatic activation (default) is enabled, this process can take up to four hours.

> **→ Tip**
>
> With large numbers of data records and for better control, use manual data activation. While automatic data activation is more convenient in many cases, it can lead to unexpected results, especially if triggered during the upload of new data records.

- Make sure to select the correct subtype when uploading the data (`supplier` for vendors or senders, and `customer` for buyers or receivers).
- Currently, products are matched by `materialNumber` only. This means that data enrichment only works for product line items that include a `materialNumber` on the document.
- If you upload a product entity without a `materialNumber`, this entity won't be matched. Always include a valid `materialNumber` when uploading product master data.
- To take advantage of ongoing normalization improvements, reupload the entire master data from time to time – for example, once a quarter. To optimize the matching of values, we make improvements of this kind continuously.

Request Examples

Not recommended – Create Enrichment Data [page 175] request payload:

```
payload:
{
    "value":[
```

```
        {
            "id":"BE0001",
            "name":"Emma Dowerg",
            "accountNumber":"SK2421",
            "address1":"Amalie-Klemm-Platz 0/9, 48581, Geithain",
            "address2":"none",        Do not add custom placeholder values
            "city":"Geithain",
            "countryCode":"DE",
            "postalCode":"48581",
            "state":"unknown",        Do not add custom placeholder values
            "email":"e.dowerf@mustermail.com",
            "phone":"",        Do not leave empty values
            "bankAccount":"DE345982837402",
            "taxId":"DE435531312"
        }
    ]
}
type: businessEntity
clientId: c_00
subtype: supplier
```

Recommended – Create Enrichment Data [page 175] request payload (do not use fields with custom placeholders or empty values):

```
payload:
{
    "value":[
        {
            "id":"BE0001",
            "name":"Emma Dowerg",
            "accountNumber":"SK2421",
            "address1":"Amalie-Klemm-Platz 0/9, 48581, Geithain",
            "city":"Geithain",
            "countryCode":"DE",
            "postalCode":"48581",
            "email":"e.dowerf@mustermail.com",
            "bankAccount":"DE345982837402",
            "taxId":"DE435531312"
        }
    ]
}
type: businessEntity
clientId: c_00
subtype: supplier
```

Related Information

14.6 Extraction Using Generative AI: Best Practices

Find out about best practices for using generative AI to extract information from documents.

> ⚠ **Restriction**
>
> Extraction using generative AI is available with the service plan **Document Information Extraction, premium edition** (`premium_edition`) only. See Service Plans [page 84] and Metering and Pricing [page 86].
>
> You can also use an SAP BTP trial account to try out extraction using generative AI. Follow the tutorial: Use Trial to Extract Information from Custom Documents with Generative AI and Document Information Extraction🏄.

> ⚠ **Caution**
>
> Bear the following in mind when using the Document Information Extraction service to process documents using generative AI:
>
> **Confidence Scores**: The Document Information Extraction service returns confidence scores for extracted results. These values are usually reliable when the service uses a pre-trained model. Be aware, however, that they can't be relied on when the service uses generative AI to extract information.
>
> **Coordinates**: Result objects returned by the API and the Document Information Extraction UI include coordinates indicating the assumed location of extracted items of information on the page. These coordinates are intended to let users see where the service extracted information and check manually for errors. Even if the extraction results are correct, some coordinates can be missing or incorrect. Therefore, coordinates can't be relied on when the service extracts information automatically using generative AI.
>
> See also Get Result [page 146] and View and Edit Extraction Results [page 252].

The better you describe the information that you want to extract using generative AI, the better your results are.

When adding fields to a schema, pay particular attention to their names and associated descriptions.

> → **Tip**
>
> When entering field names, it's often useful to imagine that you're explaining what you want to extract to a person with no prior knowledge. This approach is important when wording your schema descriptions since Document Information Extraction uses these descriptions as prompts for the LLM.

With these points in mind, we recommend the following best practices:

- Consider the wording of names and descriptions carefully, making sure that they're accurate, complete, and unambiguous.

  > ♣ **Example**
  >
  > Say you wanted the schema to extract information in a certain way from line item fields in a pharmaceutical certificate of analysis (CoA). To prompt the LLM to deliver the desired results, word the schema description as follows:

```
Extract relevant information from the following three line item fields
of a pharmaceutical CoA: Parameter, Specification, Result. Sometimes test
parameters have a title (for example, related substances) and a subtitle. If
this is the case, group the title and subtitle together and extract them as
one value.
```

- Write your definitions in English, even if documents for extraction are in a different language.
- Make sure that field names are self-explanatory and don't include abbreviations or acronyms.

> ⚙ Example
>
> Use `purchaseOrderNumber`, not `pon` or `id1`.

- If one field can have different names, include as many of these names as possible in your description.

> ⚙ Example
>
> The *Order Number* field can be called *Your Reference* in some documents.

- If there are multiple fields with similar names, add all the fields to your schema, even if only one is needed in the downstream application. Doing so simplifies processing because you can be sure of extracting a value automatically, which you can later correct manually, if necessary.

> ⚙ Example
>
> The field names *receiver material number* and *sender material number* are very similar and therefore could be confused with each other.

- Use generic terms rather than business roles in field names.
  If a document doesn't include labels indicating business roles, such as *vendor* or *customer*, there's a risk of the extraction model not knowing these roles. So, if you want to extract a vendor address from an invoice document it's best to use `senderAddress`, rather than `vendorAddress`.
- To simplify subsequent processing, make sure that your description includes the desired output format for results.

> ⚙ Example
>
> If you want a value extracted from a document to be output in uppercase, you can specify this formatting in the description.

Related Information

15 Technical Constraints

All Document Information Extraction endpoints exposed to the end user have strict technical limits. See details in the following table.

> ⓘ Note
>
> The technical limits listed here are relevant only to users of the service plans **Base Edition** (`blocks_of_100`) and **Premium Edition** (`premium_edition`) for enterprise accounts. See Service Plans [page 84].

Variable	Maximum Limit
Document file size	50 MB
Uploaded documents per hour per tenant	2000
Pages per document	100
Number of clients created per tenant	5000
Number of clients created in one API call	5000
Number of enrichment data records per tenant	100,000
Number of schemas per client	1000
Number of header fields and line items per schema	500
Number of templates per schema	1000
Number of associated documents per template	5

> ⓘ Note
>
> The Document Information Extraction service supports extraction from single or multiple tables. A single table can extend across multiple pages. It's not possible to extract information from multiple tables if they have different sets of line item fields.

> → Tip
>
> See the following sections of the Document Information Extraction documentation for other useful information:
>
> - Supported Document Types and File Formats [page 91]
> - Supported Languages and Countries/Regions [page 93]
> - Optical Character Recognition (OCR): Best Practices [page 272]

> ⚠ **Restriction**
>
> Use **only** the following types of characters for the IDs of clients, enrichment data records, system and company codes, and the name of templates, schemas, and schema header and line item fields:
>
> - letters (lowercase and uppercase)
> - numbers
> - underscore "_"
> - hyphen "-"
> - period "."
> - comma ","
> - ampersand "&"
> - dollar sign "$"
> - hashtag "#"
> - tilde "~"

Related Information

15.1 Free Tier Option and Trial Account Technical Constraints

When using the free tier option for Document Information Extraction or a trial account, be aware of the following technical limits:

> ⓘ **Note**
>
> The technical limits listed here are relevant only to users of the **Free** service plan for enterprise accounts and the Base Edition (`blocks_of_100`) service plan for trial accounts. See Service Plans [page 84].

Variable	Maximum Limit
Uploaded document pages per tenant in a rolling period of 30 days	50

> **→ Tip**
>
> The rolling period consists of the past 30 days. The total number of document pages available at any time is calculated based on how many pages you've uploaded during these 30 days.
>
> Let's say that you upload your first documents to the service on June 1, when you add 5 document pages. Up to and including June 29, you then upload 35 more pages. Because of the 50-page limit for the rolling 30-day period, you can upload only 10 more document pages on June 30.
>
> If you don't upload any pages on June 30 and wait until July 1, you can now add up to 15 more pages to the service. This is because the 5 pages you uploaded on June 1 are no longer considered since they now fall outside of the 30-day rolling period.

Variable	Maximum Limit
Pages per document	40
Number of clients created per tenant	1

> **→ Tip**
>
> A `default` client is created following tenant provisioning, enabling you to use the service immediately.

Variable	Maximum Limit
Number of enrichment data records per tenant	10
Number of schemas per client	1000
Number of header fields and line items per schema	500
Number of templates per tenant	3
Number of associated documents per template	5

> **ⓘ Note**
>
> You can't change the details of the `default` client, a previously created customized client, and enrichment data records. Delete the client and data records, and then create new ones with the updated details. For more information, see Client API [page 114] and Enrichment Data API [page 174].

See also Tutorials [page 108].

16 Extracted Header Fields

See below the list of fields that can be extracted from header fields by Document Information Extraction.

Category	Field Name	Field Label	Description	Supported Document Type	Type	Enrich-ment Data

amounts	currencyCode	Currency Code	Three-character combination codes representing each one of the world currencies in circulation. For example:
			• AUD for Australian dollar
			• CAD for Canadian dollar
			• CHF for Swiss Franc
			• EUR for euro
			• GBP for Great Britain pound (sterling)
			• USD for U.S. dollar
				`invoice`
				`payment Advice`
				`purchas eOrder`
					String	

amounts	grossAmount	Gross Amount	Invoice amount including taxes and shipping/handling costs.	`invoice`	Number	
amounts	grossAmount	Payment Amount	Amount to be paid.	`payment Advice`	Number	
amounts	grossAmount	Total Amount	Sum of subtotal, taxes, special handling charges, and shipping charges, without discounts, or total amount due and payable.	`purchas eOrder`	Number	
amounts	netAmount	Net Amount	Invoice amount without taxes and shipping/handling costs.	`invoice`	Number	
amounts	netAmount	Sub Total Amount	Amount without taxes and shipping/handling costs.	`purchas eOrder`	Number	
amounts	shippingAmount	Shipping Amount	Shipping and handling charges.	`invoice`	Number	
amounts	taxAmount	Tax Amount	The tax amount applied to this document.	`invoice`	Number	

Category	Field Name	Field Label	Description	Supported Document Type	Type	Enrichment Data
amounts	*taxId*	*Supplier Tax ID*	The number used to identify the supplier's company for tax purposes.	`invoice`	String	Used for BusinessEntity [page 178] sender and receiver enrichment.
amounts	*taxId*	*Business Partner Tax ID*	Tax identifier of the business partner sending the payment advice.	`payment Advice`	String	Used for BusinessEntity [page 178] sender and receiver enrichment.
amounts	*taxId*	*Tax ID*	Tax identifier of the buyer's business entity. Unique to each buyer. Also known as Value Added Tax (VAT) number.	`purchas eOrder`	String	Used for BusinessEntity [page 178] sender and receiver enrichment.
amounts	*taxIdNumber*	*Tax ID Number*	Tax identifier number of the buyer's business entity. Unique to each buyer. This field is specific for Germany and is not relevant for other countries.	`purchas eOrder`	String	
amounts	*taxName*	*Tax Description*	A brief description of the tax. For example: California sales tax.	`invoice`	String	
amounts	*taxRate*	*Tax Rate*	Primary tax rate applied to the document.	`invoice`	Number	
contact	*barcode*	*Barcode*	The decoded content of the QR code for business cards supports the vCard standard. Also known as VCF (Virtual Contact File), a vCard is a file format standard for electronic business cards. They can contain name and address information, phone numbers, email addresses, URLs, logos, photographs, and audio clips.	`busines sCard`	String	
contact	*buildingName*	*Building Name*	Name of the building in the address.	`busines sCard`	String	

Category	Field Name	Field Label	Description	Supported Document Type	Type	Enrich- ment Data
contact	*city*	*City*	Name of the city in the address.	busines sCard	String	
contact	*departmentN ame*	*Department*	The area one works in a company.	busines sCard	String	
contact	*email*	*Email*	Email address.	busines sCard	String	
contact	*faxNumber*	*Fax Number*	Fax phone number.	busines sCard	String	
contact	*firstName*	*First Name*	The name that stands first in one's full name.	busines sCard	String	
contact	*fixedLine*	*Fixed Line*	Landline phone number.	busines sCard	String	
contact	*houseNumber*	*House Number*	Number of the house in the address.	busines sCard	String	
contact	*lastName*	*Last Name*	Surname or family name.	busines sCard	String	
contact	*middleName*	*Middle Name*	Name between one's first name and surname.	busines sCard	String	
contact	*mobile*	*Mobile Phone*	Mobile phone number.	busines sCard	String	
contact	*namePrefix*	*Name Prefix*	Title used before a person's name.	busines sCard	String	
contact	*nameSuffix*	*Name Suffix*	Title used after a person's name.	busines sCard	String	
contact	*organizationN ame*	*Organization Name*	Company name.	busines sCard	String	
contact	*poBox*	*Post Office Box Number*	Post office box number.	busines sCard	String	
contact	*role*	*Role*	The position one has in a company.	busines sCard	String	
contact	*state*	*State*	Name of the state in the address.	busines sCard	String	
contact	*streetName*	*Street Name*	Name of the street in the address.	busines sCard	String	
contact	*website*	*Website*	Set of related web pages located under a single domain name, typically created by a single person or company.	busines sCard	String	

Category	Field Name	Field Label	Description	Supported Document Type	Type	Enrichment Data
contact	*zipCode*	*Zip Code*	Postal code of the address.	`busines sCard`	String	
details	*barcode*	*Barcode*	The decoded content of the QR code or barcode. For example: an URL (or some other text) that can be used for further processing. For more information, see Barcode Header Field in Invoice Documents [page 298].	`invoice`	String	
details	*invoiceReferenceNumber*	*Invoice Reference Number*	Number of the invoice reference. This header field is only relevant for e-invoices from India and is extracted from a barcode in the documents.	`invoice`	String	
details	*purchaseOrderNumber*	*Purchase Order*	Number of the buyer's purchase order.	`invoice`	String	
details	*quantity* ⚠ Caution We plan to deprecate the *quantity* header field in September 2024. Following that date, the field will no longer be available for extraction.	*Quantity*	Quantity of goods or services.	`purchas eOrder`	Number	
document	*documentDate*	*Invoice Date*	Date of the invoice document.	`invoice`	Date	
document	*documentDate*	*Payment Date*	Date of the payment advice document.	`payment Advice`	Date	
document	*documentDate*	*Purchase Order Date*	Date of the purchase order document.	`purchas eOrder`	Date	

Category	Field Name	Field Label	Description	Supported Document Type	Type	Enrichment Data
document	*documentNumber*	*Invoice Number*	Number that identifies this invoice.	`invoice`	String	
document	*documentNumber*	*Payment Reference*	Number of the payment advice that references the payment.	`payment Advice`	String	
document	*documentNumber*	*Purchase Order Number*	Number that identifies this purchase order.	`purchaseOrder`	String	
payment	*discount*	*Discount*	Amount deduced from gross amount.	`invoice`	String	
payment	*dueDate*	*Due Date*	Expected date of payment in extended ISO 8601 format (YYYY-MM-DD).	`invoice`	Date	
payment	*paymentTerms*	*Payment Terms*	Payment terms as found on the invoice document. Payment terms are a combination of the payment due date and the discount rate or penalty rate.	`invoice`	String	
payment	*paymentTerms*	*Payment Terms*	Indicate when payments should be made and how.	`purchaseOrder`	String	
receiver	*receiverAddress*	*Buyer Address*	Address of the organization that ordered the goods or services.	`invoice`	String	Used for BusinessEntity [page 178] receiver enrichment.
receiver	*receiverContact*	*Buyer Contact*	Name of the employee that should receive this invoice.	`invoice`	String	Used for Employee [page 180] enrichment.
receiver	*receiverId*	*Supplier ID*	A unique code that identifies the supplier.	`purchaseOrder`	String	
receiver	*receiverName*	*Buyer Name*	Name of the organization that ordered the goods or services.	`invoice`	String	Used for BusinessEntity [page 178] receiver enrichment.
receiver	*receiverTaxId*	*Buyer Tax ID*	Tax identifier of the buyer's business entity. Unique to each buyer.	`invoice`	String	

Category	Field Name	Field Label	Description	Supported Document Type	Type	Enrichment Data
sender	*senderAddress*	*Supplier Address*	Address of the organization generating this invoice.	`invoice`	String	Used for Business-sEntity [page 178] sender enrichment.
sender	*senderAddress*	*Business Partner Address*	Address of the business partner sending the payment advice.	`payment Advice`	String	Used for Business-sEntity [page 178] sender enrichment.
sender	*senderAddress*	*Buyer Address*	Address of the buyer, only one box for the street, city, and country/region of the buyer.	`purchas eOrder`	String	Used for Business-sEntity [page 178] sender enrichment.
sender	*senderBankAccount*	*Supplier Bank Account*	Bank account of the organization generating this invoice.	`invoice`	String	Used for Business-sEntity [page 178] sender and receiver enrichment.
sender	*senderBankAccount*	*Buyer Bank Account*	Bank account number of the buyer.	`purchas eOrder`	String	Used for Business-sEntity [page 178] sender and receiver enrichment.
sender	*senderCity*	*Buyer City*	City or town name of the buyer's address.	`purchas eOrder`	String	
sender	*senderCountryCode*	*Buyer Country*	Country/Region code of the buyer's address.	`purchas eOrder`	String	
sender	*senderDistrict*	*Buyer District*	District name of the buyer's address.	`purchas eOrder`	String	
sender	*senderEmail*	*Buyer Email*	Email address of the buyer.	`purchas eOrder`	String	
sender	*senderExtraAddressPart*	*Buyer Extra Address*	Any part of the buyer's address not included in the other address fields.	`purchas eOrder`	String	

Category	Field Name	Field Label	Description	Supported Document Type	Type	Enrich-ment Data
sender	*senderFax*	*Buyer Fax*	Fax number of the buyer.	purchaseOrder	String	
sender	*senderHouse Number*	*Buyer House Number*	House number of the buyer's address.	purchaseOrder	String	
sender	*senderId*	*Buyer ID*	A unique code that identifies the buyer.	purchaseOrder	String	
sender	*senderName*	*Supplier Name*	Name of organization generating this invoice.	invoice	String	Used for BusinessEntity [page 178] sender enrichment.
sender	*senderName*	*Business Partner Name*	Name of the business partner sending the payment advice.	paymentAdvice	String	Used for BusinessEntity [page 178] sender enrichment.
sender	*senderName*	*Buyer Name*	Name of the buyer.	purchaseOrder	String	Used for BusinessEntity [page 178] sender enrichment.
sender	*senderPhone*	*Buyer Phone*	Telephone number of the buyer.	purchaseOrder	String	
sender	*senderPostal Code*	*Buyer Postal Code*	Postal code of the buyer's address.	purchaseOrder	String	
sender	*senderState*	*Buyer State*	State or province name of the buyer's address.	purchaseOrder	String	
sender	*senderStreet*	*Buyer Street*	Street name of the buyer's address.	purchaseOrder	String	
shipTo	*deliveryDate*	*Delivery Date*	Date of the delivery in extended ISO 8601 format (YYYY-MM-DD).	invoice purchaseOrder	Date	
shipTo	*deliveryNoteNumber*	*Delivery Note Number*	Unique identifier on the invoice following the goods.	invoice	String	
shipTo	*shippingTerms*	*Shipping Terms*	Indicate when the goods should be delivered and how.	purchaseOrder	String	

Category	Field Name	Field Label	Description	Supported Document Type	Type	Enrich-ment Data
shipTo	*shipToAddress*	*Shipping Address*	Address where the goods will be shipped to: only one box for the street, city, and country/region.	`purchaseOrder`	String	
shipTo	*shipToCity*	*Shipping City*	City or town name of the shipping address.	`purchaseOrder`	String	
shipTo	*shipToCountryCode*	*Shipping Country*	Country/Region code of the shipping address.	`purchaseOrder`	String	
shipTo	*shipToDistrict*	*Shipping District*	District name of the shipping address.	`purchaseOrder`	String	
shipTo	*shipToEmail*	*Shipping Email*	Email address for the shipping address.	`purchaseOrder`	String	
shipTo	*shipToExtraAddressPart*	*Shipping Extra Address*	Any part of the shipping address not included in the other address fields.	`purchaseOrder`	String	
shipTo	*shipToFax*	*Shipping Fax Number*	Fax number for the shipping address.	`purchaseOrder`	String	
shipTo	*shipToHouseNumber*	*Shipping House Number*	House number of the shipping address.	`purchaseOrder`	String	
shipTo	*shipToName*	*Shipping Company Name*	Company name for the shipping address.	`purchaseOrder`	String	
shipTo	*shipToPhone*	*Shipping Telephone Number*	Telephone number for the shipping address.	`purchaseOrder`	String	
shipTo	*shipToPostalCode*	*Shipping Postal Code*	Postal code of the shipping address.	`purchaseOrder`	String	
shipTo	*shipToState*	*Shipping State*	State or province name of the shipping address.	`purchaseOrder`	String	
shipTo	*shipToStreet*	*Shipping Street*	Street name of the shipping address.	`purchaseOrder`	String	

16.1　Barcode Header Field in Invoice Documents

When the barcode header field is requested for extraction, the Document Information Extraction service scans the whole document for 1D and 2D barcodes and provides the extracted content of the barcode as a string value. The service can detect multiple barcodes in the same document and provide all the detected content in the extracted results. The most common types of 1D and 2D barcodes are supported by this field, for example:

- Code39
- Code128
- DataMatrix
- EAN
- Interleaved 2 of 5
- PDF417
- QRCode
- UPC

Document Quality and Extraction

The document quality affects the result of the extraction. For example, a low quality (low resolution) image of a scanned document with a barcode may not return any barcode header field extraction, if the barcode in the document is not identifiable. Therefore, the quality of a decoded barcode interferes in the prediction confidence score of the barcode header field. Use high quality (high resolution) images to increase the chance of extraction for barcodes in the document.

17 Extracted Line Items

See below the list of fields that can be extracted from line items by Document Information Extraction.

Category	Field Name	Field Label	Description	Supported Document Type	Type	Enrichment Data
amounts	currencyCode ⚠ Caution The currency Code line item was deprecated in February 2024. It's no longer available for extraction.	Currency Code	Three-character combination codes representing each one of the world currencies in circulation. For example: • AUD for Australian dollar • CAD for Canadian dollar • CHF for Swiss Franc • EUR for euro • GBP for Great Britain pound (sterling) • USD for U.S. dollar	purchas eOrder	String	
amounts	deductionAm ount	Deductions	Deductions for a document because of damages or late delivery.	payment Advice	Number	
amounts	discountAmo unt	Discount Amount	Discount received for a document.	payment Advice	Number	
amounts	netAmount	Amount	Total amount of the line item (typically unit price * quantity).	invoice payment Advice purchas eOrder	Number	
details	customerMat erialNumber	Customer Material Number	Unique code that identifies a specific good or service in a customer catalog or system.	purchas eOrder	String	Used for Product [page 180] enrichment.
details	description	Description	Textual description of goods or services.	invoice purchas eOrder	String	

Category	Field Name	Field Label	Description	Supported Document Type	Type	Enrich-ment Data
details	*materialNumber*	*Material Number*	Unique code that identifies a specific good or service in a supplier catalog or system.	`invoice`	String	Used for Product [page 180] enrich-ment.
details	*purchaseOrderNumber*	*Purchase Order Number*	Number of the associated purchase order (if available on line item field level).	`invoice`	String	
details	*quantity*	*Quantity*	Quantity of goods or services.	`invoice` `purchaseOrder`	Number	
details	*supplierMaterialNumber*	*Supplier Material Number*	Unique code that identifies a specific good or service in a supplier catalog or system.	`purchaseOrder`	String	Used for Product [page 180] enrich-ment.
details	*unitOfMeasure*	*Unit of Measure*	The unit of measure UN/CEFACT code. For example: EA for each, HR for hour and YR for year.	`invoice` `purchaseOrder`	String	
details	*unitPrice*	*Unit Price*	Price for a single instance of an object.	`invoice` `purchaseOrder`	Number	
document	*documentDate*	*Document Date*	Date of the invoice document.	`payment Advice`	Date	
document	*documentDate*	*Document Date*	Requested delivery date.	`purchaseOrder`	Date	
document	*documentNumber*	*Document Number*	Document number that is used by the receiver.	`payment Advice`	String	
item	*itemNumber*	*Item Number*	Item number that is used by the receiver.	`purchaseOrder`	String	

18 Security

Get an overview on the security information that applies to Document Information Extraction. Learn about the main security aspects of the service and its components.

Related Information

18.1 Data Protection and Privacy

Introduction

Data protection is associated with numerous legal requirements and privacy concerns. In addition to compliance with general data privacy regulation, it is necessary to consider compliance with industry-specific legislation in different countries/regions. SAP provides specific features and functions to support compliance with regard to relevant legal requirements, including data protection. SAP does not give any advice on whether these features and functions are the best method to support company, industry, regional, or country/region-specific requirements. Furthermore, this information does not give any advice or recommendation in regards to additional features that would be required in particular IT environments; decisions related to data protection must be made on a case-by-case basis, under consideration of the given system landscape and the applicable legal requirements.

> ⓘ Note
>
> SAP software supports data protection by providing security features and specific data protection-relevant functions such as functions for the simplified blocking and deletion of personal data. SAP does not provide legal advice in any form. The definitions and other terms used in this document are not taken from any given legal source.

Document Information Extraction may process personal data, such as employee names and email addresses, depending on the information available in documents and enrichment data.

All data processed by the service is stored in the SAP BTP, Cloud Foundry environment. Document Information Extraction generally processes the following data types:

Data required by Document Information Extraction

Data	Purpose
Inference Documents	Refers to documents that are submitted by users to receive machine learning predictions.
Data Feedback Collection Documents	Refers to documents that are submitted by users to receive machine learning predictions, and to be used to retrain the service's machine learning models through the data feedback collection feature.
Documents Associated with Templates	Refers to documents that are submitted by users and associated with templates to extract information from other similar business documents.
Enrichment Data	Refers to enrichment data records, for example, supplier name and supplier address. The service matches your existing structured data (typically master data records) with the information extracted from documents.

> ⓘ Note
>
> Document Information Extraction does not provide the technical capabilities to support the collection, processing, and storage of sensitive personal data.

Read Access Logging

The data used by Document Information Extraction is controlled and managed by the consuming application or customer that uses the service. However, the service does not have any means to verify whether the data uploaded to the service contains any sensitive personal data. Therefore, Document Information Extraction does not support logging of read access to sensitive personal data. For more information, see Auditing and Logging Information [page 305].

Information Report

The data from inference documents and data feedback collection documents used by Document Information Extraction is controlled and managed by the consuming application which calls the Document Information Extraction APIs. Document Information Extraction does not create or modify inference or retraining data provided by the consuming application. Therefore it is not possible for Document Information Extraction to provide a retrieval function to identify data of specific individuals.

It is recommended that the consuming application which uses Document Information Extraction provides personal data reports to its users and transfers to Document Information Extraction for processing. After every change of the data in the customer system, customers should call the Create Enrichment Data [page 175] endpoint.

Deletion of Personal Data

See in the table below, retention period and deletion details for all data types required by the Document Information Extraction service.

Deletion of personal data is logged using audit logging services. For more information, see Audit Logging in the Cloud Foundry Environment.

Data	Deletion
Inference Documents	The default retention period for inference data documents is 7 days. You can also use the `documentRetentionTimeDays` key to Create Configuration [page 122] and customize the retention period, for inference documents uploaded to the service, from 1 to 30 days. You can delete inference data using the Delete Document [page 174] endpoint at any time, even before the retention period expires.
Data Feedback Collection Documents	There is no default retention period for retraining data documents. You can delete all retraining data using the Create Configuration [page 122] and Delete Configuration [page 131] endpoints at any time. You can also individually delete documents previously submitted for retraining using the Delete Document [page 174] endpoint at any time. If the `performPIICheck` subconfiguration is set to *true*, the service automatically scans all submitted documents and tries to exclude all documents where Personally Identifiable Information (PII) data is detected from being used for retraining and improving the service. It is the customer's responsibility to ensure that no personal data is submitted when using the data feedback collection feature.
Documents Associated with Templates	The documents uploaded to the document feature and associated with templates are not deleted automatically. To minimize the processing of personal data, do not use sample documents that contain personal data.
Enrichment Data	Enrichment data containing personal data is deleted automatically when customers delete the service instances. You also control the enrichment data retention period using the Delete Enrichment Data (Synchronous) - Deprecated [page 189] and Delete Enrichment Data (Asynchronous) [page 191] endpoints to delete enrichment data records at any point in time.

Change Log

The application does not perform any update of enrichment data automatically. Any update of enrichment data per customer request would be logged using audit logging services. For more information, see Audit Logging in the Cloud Foundry Environment.

Consent

According to Personal Data Processing Agreement for SAP Cloud Services, SAP acts as data processor. Thus, customers are responsible for obtaining relevant consent to process personal data, including when applicable approval by controllers to use SAP as a processor.

18.2 Auditing and Logging Information

Here you can find a list of the security events that are logged by the Document Information Extraction service.

Security events written in audit logs

Event grouping	What events are logged	How to identify related log events	Additional information
Authentication related events	Authentication success	Successful login attempt for tenant {tenant_id} on {instance_id} on {time}	See below the definitions of the notations used in the log events.
	Authentication failure	Failed login attempt for tenant {tenant_id} on {instance_id} on {time}	• {client_name}: ID of a client created with the Create Client [page 114] endpoint.
Client related events	Client(s) created	"Tenant" and ID consisting of: targetTenant {tenant_id} (Multiple) Attribute(s) with name "client {client_name}" and state change from None to "CREATED"	• {dataset_id}: ID of the dataset (enrichment data). • {document_id}: ID of a document uploaded to the service.
	Client(s) deleted	"Tenant" and ID consisting of: targetTenant {tenant_id} (Multiple) Attribute(s) with name "client {client_name}" and state change from "CREATED" to "DELETED"	• {instance_id}: ID of the service instance used to access the service. • {tenant_id}: ID of the tenant used to access the service.
Dataset related events	Modification of dataset (enrichment data)	Modification of dataset:{dataset_id} successful Modification of dataset:{dataset_id} failed	• {time}: time stamp of when a log was created. You can use time stamps to sort the logs by time. • {user_id}: ID of the user that accessed the service and performed document related tasks.

Event grouping	What events are logged	How to identify related log events	Additional information
	Deletion of dataset (enrichment data)	Deletion of dataset:{dataset_id} successful	
		Deletion of dataset:{dataset_id} failed	
Document related events	Deletion of documents (customer documents, for example, invoices uploaded to the service)	Deletion of document:{document_id} successful	
		Deletion of document:{document_id} failed	
	Document access attempt	Document access attempt by {user_id} of {tenant_id} on {document_id}	
	Document updated	Successful / failed modification attempt by {user_id} of {tenant_id} on {document_id}	
		Attribute with name "extractions" was changed	
	Document confirmed	Successful / failed modification attempt by {user_id} of {tenant_id} on {document_id}	
		Attribute with name "status" was changed to "CONFIRMED"	
	Document deleted	Successful / failed modification attempt by {user_id} of {tenant_id} on {document_id}	
		Attribute with name "status" was changed to "DELETED"	
Tenant related events	Tenant provision	"Tenant provisioned" and ID consisting of: targetTenant {tenant_id}	
		Attribute with name "state" was changed from "DOES_NOT_EXIST" to "PROVISIONED"	

Event grouping	What events are logged	How to identify related log events	Additional information
	Tenant de-provision	"Tenant de-provisioned" and ID consisting of: targetTenant {tenant_id}	
		Attribute with name "state" was changed from "PRO-VISIONED" to "DEPROVI-SIONED"	
	Tenant saas-subscription	"Tenant SAAS Subscription" and ID as targetTenant {tenant_id}	
		Attribute with name "state" was changed from "DOES_NOT_EXIST" to "SAAS_SUBSCRIBED"	
	Tenant saas-unsubscription	"Tenant SAAS UnSubscription" and ID as targetTenant {tenant_id}	
		Attribute with name "state" was changed from "SAAS_SUBSCRIBED" to "SAAS_UNSUBSCRIBED"	

Related Information

Audit Logging in the Cloud Foundry Environment

18.3 Front-End Security

The Document Information Extraction UI (User Interface) is a web application that supports the following features:

- SAPUI5 Frame option to avoid clickjacking attacks
- Cross-site request forgery (CSRF) protection
- Cross-site scripting (XSS) output encoding during SAP UI5 rendering
- Secure transport layer security (TLS) encryption using HTTPS

- Access to business data only after authentication and with sufficient authorizations using SAP Business Technology Platform (SAP BTP) identity management and SAP BTP role-based access management (RBAM)
- Cross-site-scripting counter measures
- Session inactivity timeout (15 minutes)
- Rate limiting for document upload
- Data access audit log for viewing extracted document
- Data change audit log for changing/confirming extraction results
- Data change audit log for deleting a document

19 Accessibility Features in Document Information Extraction

To optimize your experience of Document Information Extraction, SAP Business Technology Platform (SAP BTP) provides features and settings that help you use the software efficiently.

> ⓘ Note
>
> Document Information Extraction runs on the SAP BTP cockpit. For this reason, the accessibility features for SAP BTP cockpit apply. For more information, see the accessibility documentation for SAP BTP cockpit on SAP Help Portal at Accessibility Features in SAP BTP Cockpit.

The Document Information Extraction UI is based on SAPUI5. It provides accessibility support in its tools and customer documentation. For more information on keyboard handling for SAPUI5 UI elements and screen-reader support for SAPUI5 controls, see Accessibility for End Users.

20 Monitoring and Troubleshooting

Find out how to get support, and explore solutions to potential issues.

Related Information

20.1 Getting Support

If you encounter an issue with this service, we recommend that you follow the procedure below.

Check Platform Status

Check the availability of the platform at SAP Trust Center .

For more information about selected platform incidents, see Root Cause Analyses.

Check Guided Answers

In the SAP Support Portal, check the Guided Answers section for SAP Business Technology Platform. You can find solutions for general platform issues as well as for specific services there.

Contact SAP Support

You can report an incident or error through the SAP Support Portal. For more information, see Getting Support.

Please use the following component for your incident:

Component Name	Component Description
CA-ML-BDP	Services related to Business Document Processing

When submitting the incident, we recommend including the following information:

- Region information (Canary, EU10, US10, for example)
- Subaccount technical name

- The URL of the page where the incident or error occurs
- The steps or clicks used to replicate the error
- Screenshots, videos, or the code entered
- Any business documents (for example, invoices) with which there have been extraction issues

Related Information

Built-In Support [page 249]

20.2 Troubleshooting

In this section, see possible reasons for the following Document Information Extraction potential issues:

20.2.1 Problem: You Receive Status Code 4**

If you are getting a 4** status code for your request (such as `400`, `401`, or `422`), make sure that you are submitting the request correctly. In most cases, the problem can be fixed in the request. Perhaps the authentication information is missing or the request is using the wrong HTTP method (GET, POST, DELETE). Or maybe the payload is invalid.

20.2.2 Problem: You Receive Status Code 400

<⟩ Output Code

```
Status: 400 Bad Request
{
"errors": [
{
"code": "string",
"message": "string"
}
]
}
```

Possible reasons:

A `400` error means that the request is malformed. This can be because of one of the following reasons:

- The request does not have the correct `Content-Type` header (usually `application/json`)
- The request payload is not a valid JSON
- The request payload does not contain some of the required fields and files
- The authorization token was not included in the headers. The error message will be `"Authorization token was not found in headers"`. The header should look like `Authorization: Bearer eyJhbGc....`

20.2.3 Problem: You Receive Status Code 401

<⟩ Output Code

```
Status: 401 Unauthorized
{
  "error": {
    "statusCode": 401,
    "message": "..."
  },
}
```

Possible reasons:

A `401` error means that you did not supply correct authentication information. This can be because of one of the following reasons:

- You provided an invalid tenant password
- You provided an invalid authentication token or the authentication token has expired

20.2.4 Problem: You Receive Status Code 413

‹›Output Code

```
Status: 413 Request Entity Too Large
{
  "error": {
    "statusCode": 413,
    "message": "..."
  }
}
```

Possible reasons:

A `413` status indicates that the request you are making is too large. Either you are sending a file that is too large or trying to process too many objects in a single request.

20.2.5 Problem: You Receive Status Code 415

‹›Output Code

```
Status: 415 Unsupported File Type
{
  "error": {
    "statusCode": 415,
    "message": "..."
  }
}
```

Possible reasons:

You get a `415` status code when you use the wrong content type or file format. See Supported Document Types and File Formats [page 91].

20.2.6 Problem: You Receive Status Code 422

```
Status: 422 Unprocessable Entity
{
  "error": {
    "statusCode": 422,
    "message": "..."
  }
}
```

Possible reasons:

You get a `422` status code when your request payload references a clientId, senderId, or documentId that does not exist. For example, you will get this error if you try to create a document for a client that does not exist.

You may also get this error if the document you upload cannot be parsed.

20.2.7 Problem: You Receive Status Code 429

```
Status: 429 Rate Limit Exceeded
{
  "error": {
    "statusCode": 429,
    "message": "..."
  }
}
```

Possible reasons:

You get a `429` status code when you have reached the rate limit for this user. You have made too many requests.

20.2.8 Problem: You Receive Status Code 500

You get a 500 status code for your request due to a server error and not an issue with the request. A `500` error is usually an error in the Document Information Extraction application code. To report `500` errors, create an incident on the component CA-ML-BDP, as described in Getting Support [page 310].

Important Disclaimers and Legal Information

Hyperlinks

Some links are classified by an icon and/or a mouseover text. These links provide additional information.
About the icons:

- Links with the icon 🡕 : You are entering a Web site that is not hosted by SAP. By using such links, you agree (unless expressly stated otherwise in your agreements with SAP) to this:

    - The content of the linked-to site is not SAP documentation. You may not infer any product claims against SAP based on this information.

    - SAP does not agree or disagree with the content on the linked-to site, nor does SAP warrant the availability and correctness. SAP shall not be liable for any damages caused by the use of such content unless damages have been caused by SAP's gross negligence or willful misconduct.

- Links with the icon 🕊 : You are leaving the documentation for that particular SAP product or service and are entering an SAP-hosted Web site. By using such links, you agree that (unless expressly stated otherwise in your agreements with SAP) you may not infer any product claims against SAP based on this information.

Videos Hosted on External Platforms

Some videos may point to third-party video hosting platforms. SAP cannot guarantee the future availability of videos stored on these platforms. Furthermore, any advertisements or other content hosted on these platforms (for example, suggested videos or by navigating to other videos hosted on the same site), are not within the control or responsibility of SAP.

Beta and Other Experimental Features

Experimental features are not part of the officially delivered scope that SAP guarantees for future releases. This means that experimental features may be changed by SAP at any time for any reason without notice. Experimental features are not for productive use. You may not demonstrate, test, examine, evaluate or otherwise use the experimental features in a live operating environment or with data that has not been sufficiently backed up.
The purpose of experimental features is to get feedback early on, allowing customers and partners to influence the future product accordingly. By providing your feedback (e.g. in the SAP Community), you accept that intellectual property rights of the contributions or derivative works shall remain the exclusive property of SAP.

Example Code

Any software coding and/or code snippets are examples. They are not for productive use. The example code is only intended to better explain and visualize the syntax and phrasing rules. SAP does not warrant the correctness and completeness of the example code. SAP shall not be liable for errors or damages caused by the use of example code unless damages have been caused by SAP's gross negligence or willful misconduct.

Bias-Free Language

SAP supports a culture of diversity and inclusion. Whenever possible, we use unbiased language in our documentation to refer to people of all cultures, ethnicities, genders, and abilities.

www.sap.com/contactsap

THE BEST RUN SAP