

Exercise 1: Document Information Extraction Premium Edition

Step 1: Access your BTP Cockpit

In your web browser, open the [SAP BTP cockpit](#) and navigate to the subaccount in which you have activated Document Information Extraction Service.

Step 2: Open DOX

In your BTP Subaccount where Document Information Extraction instance has been instantiated, click on “Instances and Subscriptions” and go to DOX application by clicking over Document Information Extraction or in the “Go to Application” button

Subscriptions (8) Instances (2) Environments (1)

Applications to which your subaccount is currently subscribed

Application	Plan	Changed On	Status	
SAP AI Launchpad	standard	11 Mar 2024	Subscribed	>
SAP Build Code	standard	9 Aug 2024	Subscribed	>
Continuous Integration & Delivery	default	30 Jul 2024	Subscribed	>
Document Information Extraction	default	27 May 2024	Subscribed	>
SAP HANA Cloud	tools	18 Mar 2024	Subscribed	>

Step 3: Create a Schema

3.1

Navigate to the Schema Configuration page and click on the Create button.

Schemas (5)

Schema Configuration delete SAP schemas.

Search Create Delete ↓↑ ▼

<input type="checkbox"/> Schema Name	Document Type	Description	SAP Schema	Status	
<input type="checkbox"/> resume_i044067	Custom	Resume fetch	NO	ACTIVE	>
<input type="checkbox"/> SAP_purchaseOrder_schema	Purchase Order	Schema configuration for document type :purchaseOrder	YES	ACTIVE	>
<input type="checkbox"/> SAP_paymentAdvice_schema	Payment Advice	Schema configuration for document type :paymentAdvice	YES	ACTIVE	>
<input type="checkbox"/> SAP_invoice_schema	Invoice	Schema configuration for document type :invoice	YES	ACTIVE	>
<input type="checkbox"/> SAP_OCROnly_schema	Custom	OCR Schema configuration for document type :custom	YES	ACTIVE	>

3.2

For this example, we will be extracting information from a resume document, so give a name in the format “resume_<your name>”. Fill in the rest of the information as shown below:

Create New Schema

Name: * resume_i123456

Description: Resume fetch

Document Type: * Custom

OCR Engine Type: * Document

Create

Cancel

3.3

Open the schema you just created.

Schema Name	Document Type	Description	SAP Schema	Status
<input checked="" type="checkbox"/> resume_i123456	Custom	Resume fetch	NO	INACTIVE
<input type="checkbox"/> resume_i044067	Custom	Resume fetch	NO	ACTIVE
<input type="checkbox"/> SAP_purchaseOrder_schema	Purchase Order	Schema configuration for document type :purchaseOrder	YES	ACTIVE
<input type="checkbox"/> SAP_paymentAdvice_schema	Payment Advice	Schema configuration for document type :paymentAdvice	YES	ACTIVE
<input type="checkbox"/> SAP_invoice_schema	Invoice	Schema configuration for document type :invoice	YES	ACTIVE
<input type="checkbox"/> SAP_OCROnly_schema	Custom	OCR Schema configuration for document type :custom	YES	ACTIVE

3.4

Click on the Add button to add a new data entry we will be looking for in the documents.

Schemas (6)

You can't delete SAP schemas.

Search

Schema Name	Status
<input type="checkbox"/> resume_i123456	INACTIVE
<input type="checkbox"/> resume_i044067	ACTIVE
<input type="checkbox"/> SAP_purchaseOrder_schema	ACTIVE
<input type="checkbox"/> SAP_paymentAdvice_schema	ACTIVE
<input type="checkbox"/> SAP_invoice_schema	ACTIVE
<input type="checkbox"/> SAP_OCROnly_schema	ACTIVE

resume_i123456

Activate Edit Copy Delete

Description: Resume fetch
Document Type: Custom
Creation Date: July 3, 2024
Status: INACTIVE

Header Fields

Field Name	Description	Data Type	Action
No data			

Line Item Fields

Field Name	Description	Data Type	Action
No data			

3.5

Add the following data entries for the header fields. Keep in mind, for simple ones like first and last name, we can count on the underlying AI to understand them without additional context, but for more complex ones like degree for example, we might want to provide some additional description of what we are looking for. You can use the table below to copy the data or refer to the images.

Field name	Description
firstName	
lastName	
degree	Highest degree or highest education
employer	Current employer or latest work experience
jobTitle	Current job title or latest work experience
techSkills	List all tech skills / stack comma separated format

Field Name: *

firstName

Field Label:

Description:

Data Type: *

string

Setup Type: *

auto

Field Name: *

lastName

Field Label:

Description:

Data Type: *

string

Setup Type: *

auto

Field Name: *

degree

Field Label:

Description:

Highest degree or highest education

Data Type: *

string

Setup Type: *

auto

Field Name: *

employer

Field Label:

Description:

Current employer or latest work experience

Data Type: *

string

Setup Type: *

auto

Field Name: *

Field Label:

Description:

Data Type: *

Setup Type: *

Field Name: *

Field Label:

Description:

Data Type: *

Setup Type: *

This is the table you should end up with after adding all header fields:

Field Name	Description	Data Type	Action
firstName		string	
lastName		string	
degree	Highest degree or highest education	string	
employer	Current employer or latest work experience	string	
jobTitle	Current job title or latest work experience	string	
techSkills	List all tech skills / stack comma separated format	string	

3.6

Now we need to add entries to the line items table as well.

Line Item Fields

Field Name

Description

Data Type

Action

No data

Repeat the same process as before for these.

Field name	Description
workExpEmployer	Employer of particular experience
workExpJobTitle	Job title of particular experience
workExpStartDate	Take the first day of the particular month for the particular experience
workExpEndDate	Take the first day of the particular month for the particular experience

Field Name: *

Field Label:

Description:

Data Type: *

string

Setup Type: *

auto

Field Name: *

Field Label:

Description:

Data Type: *

string

Setup Type: *

auto

Field Name: *

Field Label:

Description:

Data Type: *

date

Setup Type: *

auto

Field Name: *

workExpEndDate

Field Label:

Description:

Take the first day of the particular month for the particular experience.

Data Type: *

date

Setup Type: *

auto

This is the table you should end up with after adding all line-item fields:

Line Item Fields

Add

Field Name	Description	Data Type	Action
workExpEmployer	Employer of particular experience	string	 
workExpJobTitle	Job title of particular experience	string	 
workExpStartDate	Take the first day of the particular month for the particular experience	date	 
workExpEndDate	Take the first day of the particular month for the particular experience	date	 

Step 4: Test the Schema

4.1

Activate the schema by using the button in the top toolbar.


resume_j123456


Activate

Edit

Copy

Delete







Description: Resume fetch

Document Type: Custom

Creation Date: July 3, 2024

Status: DRAFT

Header Fields

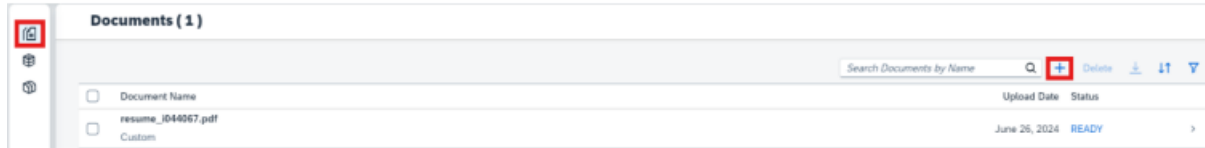
Line Item Fields

4.2

Download the resume sample document shared in the meeting.

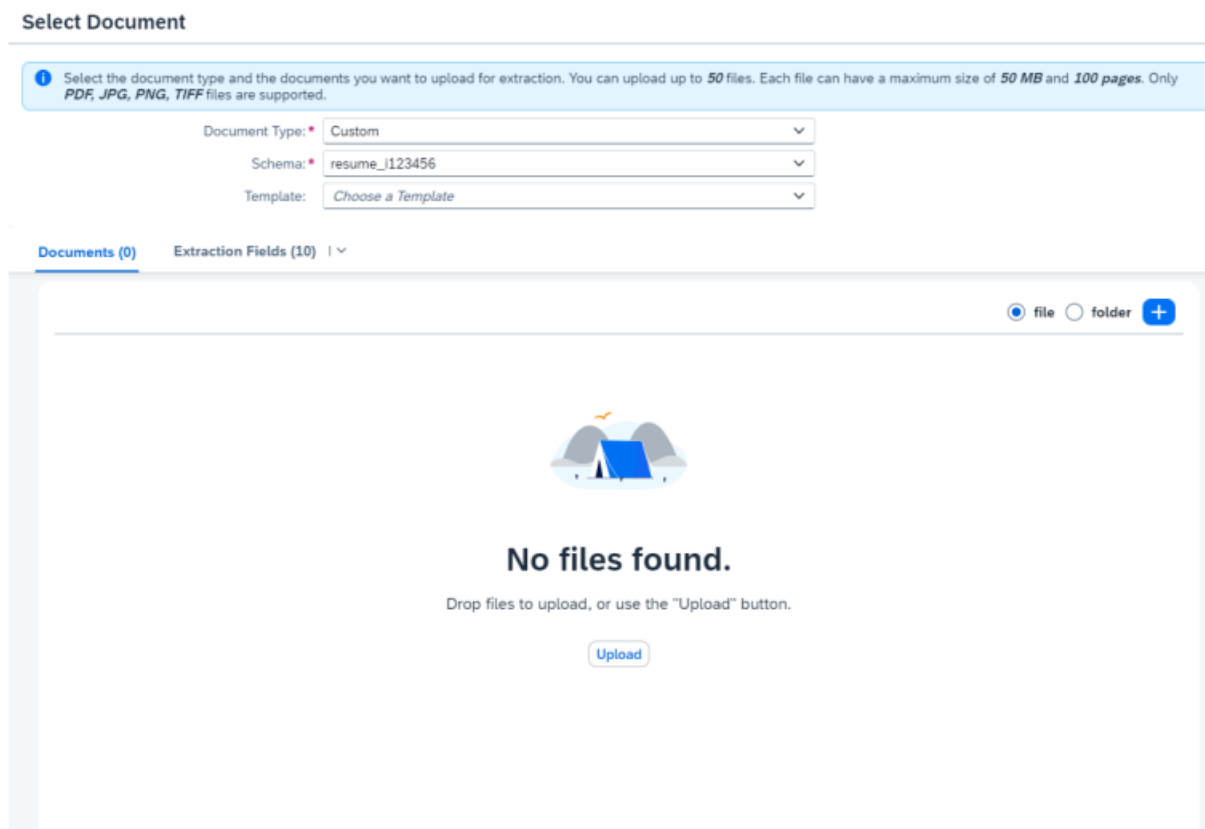
4.3

Navigate to the Document page and click on the plus button to add a new document.



4.4

Select the Custom type and our newly created Schema. Do not select a template. Click the Upload button or drop the resume file in the upload box.



4.5

Click on the Confirm button.

Clicking on Extraction Results will also give you a table with the data-field matching.

resume.pdf

Edit



Extraction Confidence Range:

↓ 0% - 50%

↗ 51% - 79%

↗ 80% - 100%

Header Fields (6)



Label	Value
degree	Bachelor of Technol...
employer	Tech Solutions Indi...
firstName	Priya
jobTitle	Senior Software En...
lastName	Sharma
techSkills	Database Manage...

Line Items (2)



Label	Value
▼ Line Item 1	
workExpEmployer	Tech Solutions Indi...
wokrExpJobTitle	Senior Software En...
workExpStartDate	2014-06-01
workExpEndDate	
▼ Line Item 2	
workExpEmployer	Infosys Technologie...
wokrExpJobTitle	Software Developer...
workExpStartDate	2013-05-01
workExpEndDate	2013-08-01

Step 5: Explore

5.1

Now you can try to use another file → find a template from the internet or upload your own resume and see the result!

5.2

Try changing the field names and the field descriptions and see how that affects the extraction. For example, take some of the more complex fields like the start and end dates for previous experiences and remove their description. You could also experiment with having just one field for name rather than separate ones.

5.3

If you want to explore the service further, try to create a schema for a different type of document. Here are some suggestions:

- Passport
- National ID
- Sales receipt
- News or science article (Challenge)

Remember that some field such as first and last name might be easy to grasp for the model, but others might require some more sophisticated description for good results.