

OCI Document AI Scanner

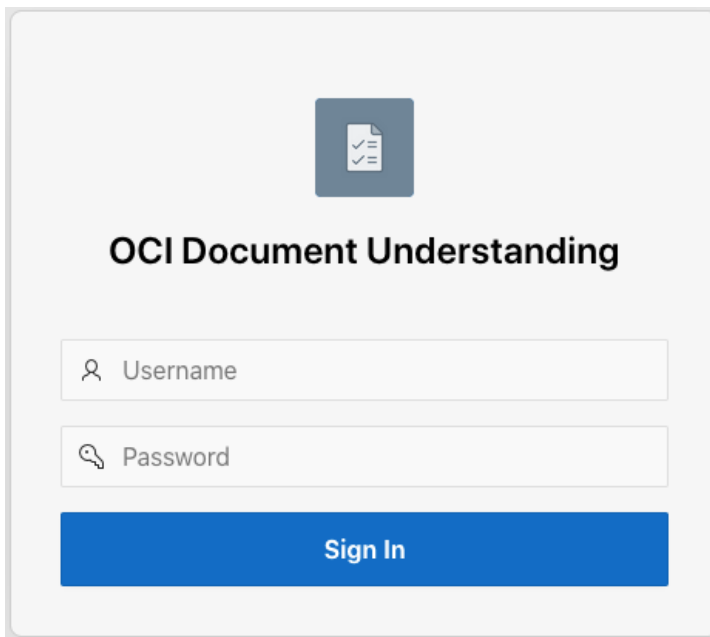
The application opens up the functions of the Oracle Cloud - Document Understanding AI Service in a modern APEX user interface. The aim of the application is to process several pdf or image files in one job and to provide the results of the analysis (text extraction, table extraction, document classification, key value extraction) as a ZIP file for download or export to an accounting database. The details can be displayed and improved on the Processor job details and Document analysis pages.

Inhaltverzeichnis

Sign up on the login page.....	3
User interface language.....	3
UI Appearance.....	4
Create processor orders.....	5
Process documents.....	6
Download results.....	6
Evaluate results.....	6
Order delete.....	7
Processor job details.....	8
Export invoices.....	9
Document analysis.....	10
document attributes.....	11
field labels.....	12
line item.....	14
invoice items.....	15
Tables.....	16
Lines of text.....	17
Text preview.....	17
Search documents.....	18
Configuration.....	19
AI processor configuration.....	19
Configuration - Field Types.....	23
Configuration - Line Item Types.....	24
Configuration - Languages.....	25
Configuration - Text Index Status.....	26
Installation.....	26
Requirements.....	26
OCI configuration.....	27
Install APEX application.....	29
Processing of processor orders.....	30
1. Store files in the object store.....	30
2. Compile and start the processor job.....	30
3. Awaiting processing.....	30
4. Processing of the results.....	30
5. Algorithm for Analyzing Text Fragments.....	31

Sign up on the login page

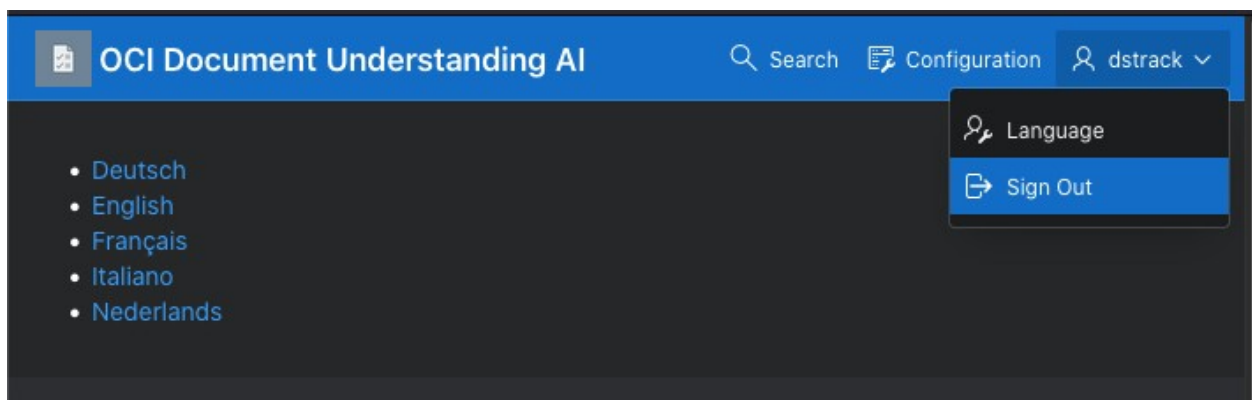
The application requires the username and password of the APEX account set up by the administrator of the APEX workspace.



The login page for OCI Document Understanding. It features a document icon with a checklist at the top. Below the icon is the title "OCI Document Understanding". There are two input fields: "Username" and "Password", each with a magnifying glass icon. A blue "Sign In" button is located below the password field.

User interface language

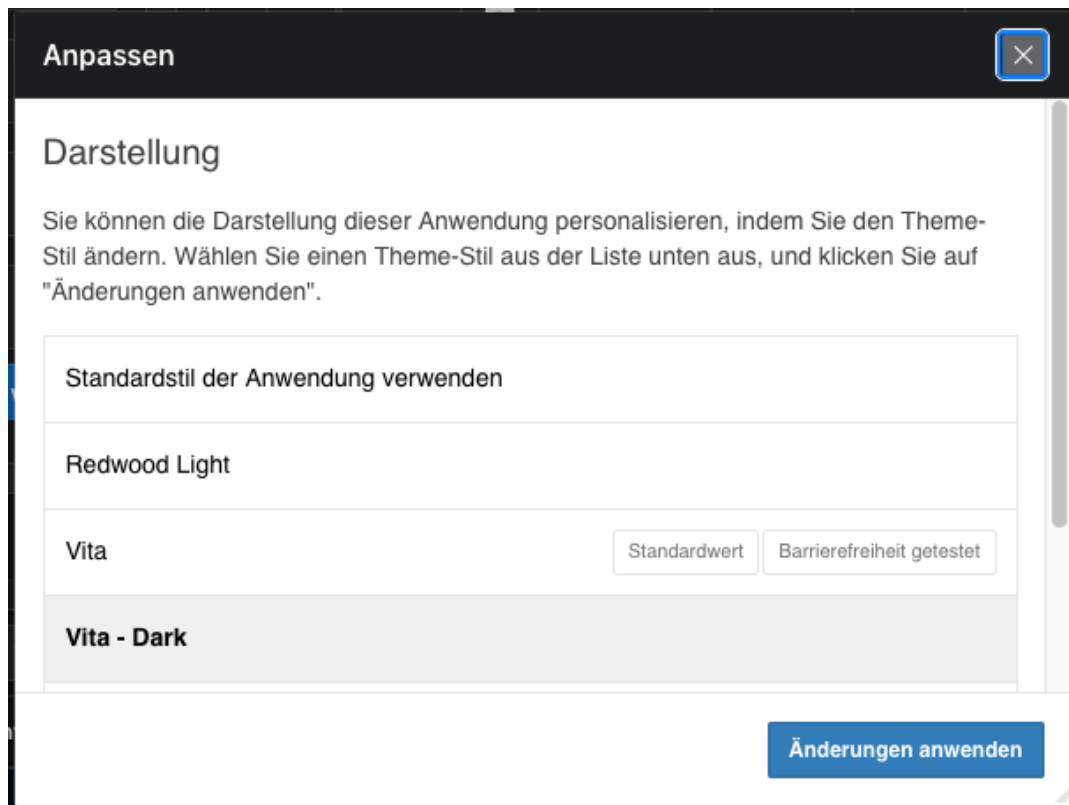
After logging in, the user interface language can be changed. To do this, click on the user icon with their name in the top right corner and select Language. A list of languages appears in the top left. Select the appropriate language by clicking on an entry. The languages German, English, French, Italian and Dutch are currently supported.



UI Appearance

Click Customize on the bottom left and a dialog box titled Appearance will appear. You can personalize the appearance of this application by changing the theme style.

Choose a theme style from the list below and click Apply Changes.



Create processor jobs

New orders can be created and completed orders can be viewed on the home page. For a new job, load multiple files by dragging and dropping them into the designated zone at the top left of the page. Set the document type, language and context of the documents.

For the import of invoices, set the document type to **Invoice**, language to **English**, and the context to the appropriate project name.

With the language setting English, it is ensured that American formatted numbers and date values can be converted, even if they occur in German documents.

With the document type setting Invoice, it is ensured that the appropriate field names are searched for, even if a different document type was determined by the document analysis.

Only single invoices per PDF file

The scans of the invoices should contain only a single invoice. The program CANNOT recognize and break down multiple invoices in one PDF document! On a Mac, with the Preview program, a new document per page can be created by drag and drop of a page onto the desktop, if necessary.

OCI Document Understanding AI

Search Configuration dstrack

Home \

Document AI Scanner

Process New Documents

Drag and Drop Files

Select files or drop in this region.

Choose Files

Document Type

Invoice

Language

English

Context

KAP SixSenses

Table Extraction

Generate Searchable Pdf

Process Documents

AI Processor Jobs

Actions Edit Save

Reset

	Job Details	Downl	Lifecyclestate	Document Ty	Language	Context	Files	Last Update Date	Last Update	Docume	Key Values	Line It	Has Been E
	000176		SUCCEEDED	Invoice	English	KAP SixSenses	WF_Maklerprovision_Re_202...	03-NOV-2023 13:04	DSTRACK	1	13	3	N
	000175		SUCCEEDED	Invoice	German	KAP SixSenses	Lindner_AR_400166.pdf	13-OCT-2023 13:14	DSTRACK	1	61	425	Y
	000174		SUCCEEDED	Invoice	German	KAP SixSenses	Webmedia_Huber_Re22013...	12-OCT-2023 15:09	DSTRACK	5	92	25	N
	000172		SUCCEEDED	Invoice	English	KAP SixSenses	Oracle_Re_300002332.pdf, ...	12-OCT-2023 11:49	DSTRACK	2	42	11	N
	000171		SUCCEEDED	Invoice	English	KAP SixSenses	Stadtgem_Mitterall_Grundst...	06-OCT-2023 13:00	DSTRACK	12	197	96	N
	000170		SUCCEEDED	Invoice	German	KAP SixSenses	Notar_Zimmerer_Re_1334_2...	06-OCT-2023 12:57	DSTRACK	12	276	310	N
	000169		SUCCEEDED	Invoice	English	KAP SixSenses	230309_Octopus.pdf, 2303...	06-OCT-2023 12:53	DSTRACK	13	301	292	N

1 rows selected

Total 7

Documents

Search: Full Text

Go

Actions

Reset

Document	File Name	Language	Language Score	Document Type	Document Type Score	Pages	Line Items	Key Values	Tables
	Oracle_Re_300002332.pdf	German	75.4%	Others	99.0%	1	5	21	0
	Oracle_Re_300006785.pdf	German	75.2%	Others	51.3%	1	6	21	0

1 rows selected

Total 2

4 von 33

Document type

Select the type of documents uploaded. The document type determines which type of field content and line item is generated in the report.

Language

Select the language of the uploaded documents. The language determines which country-specific data formats are used when converting to numbers and dates.

Context

Select the context for the uploaded documents. In the case that the results after the analysis e.g. are to be assigned to a project, you can set the project here.

Table extraction

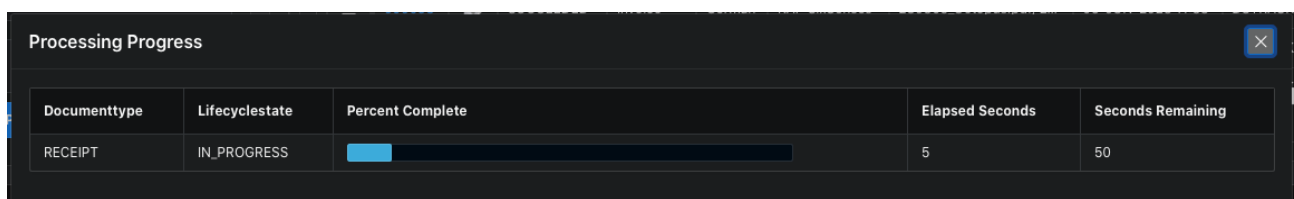
Activate this option if you want to evaluate the tables contained in the documents. Table extraction can only be used with some document types.

Create searchable PDF

Activate this option if you want to convert the uploaded documents into searchable PDF. This makes these documents searchable with an Oracle Text Index.

Process documents

Start the job by clicking on 'Process documents'. A dialog window informs about the progress and the completed percentage of the job during execution.

A dark-themed dialog box titled "Processing Progress" with a close button in the top right corner. It contains a table with five columns: Documenttype, Lifecyclestate, Percent Complete, Elapsed Seconds, and Seconds Remaining. The first row shows "RECEIPT" for Documenttype, "IN_PROGRESS" for Lifecyclestate, a progress bar at approximately 10% completion for Percent Complete, "5" for Elapsed Seconds, and "50" for Seconds Remaining.

Documenttype	Lifecyclestate	Percent Complete	Elapsed Seconds	Seconds Remaining
RECEIPT	IN_PROGRESS	<div><div></div></div>	5	50

When the job is complete, the dialog box will disappear and the **processor job details** will be displayed.

Download results

Click on the ZIP file icon in the Download column of the AI Processor Jobs list to start the download. The zip file contains a directory structure with the various lists of results. The folder has the name Document_AI_Files_X and ends with the job number.

The directories csv_files and searchable_pdf are included.

In csv_files you will find the files that you can evaluate in Numbers or Excel. Depending on the type of order, there are files with the names of the documents and the extensions X_Line_Items.csv

and X_Tables.csv. Also Invoices_List_X.csv, Receipts_List_X.csv, Driver_License_List_X.csv, Passport_List_X.csv files with field values for each corresponding document.

The searchable_pdf directory contains the searchable PDF files for the uploaded documents.

Evaluate results

Select an entry in the AI Processor Jobs list by clicking on the desired row. The individual documents for the selected entry are listed in the Documents list. Click on the job number in the Job Details column to view processor job details, or click on the magnifying glass icon in the Document column to view the document analysis results.

Delete Job

After you have downloaded or evaluated the results, the job can be deleted. To do this, select 'Delete row' from the row menu and then click on Save. With the deletion, the memory used in the database and also in the object store is released again.

Processor job details

This page can be used to view order identification data and a quick view of recognized field labels with content for invoices with invoice line items and receipts with receipt line items. Click on the magnifying glass icon in the Document column to open the document analysis.

OCI Document Understanding AI

Search Configuration dstrack

Home

Processor Job Details

PreviousAssign Key AliasesExport To AccountingDownload ZIP FileProcess AnotherNext

Job ID
000172

Creation Date
12-OCT-2023 11:49

Created By
DSTRACK

Lifecyclestate
SUCCEEDED

Lifecycledetails

Processorjob Message

Context
KAP SixSenses

Language Code
English

Documenttype
Invoice

Documents
2

Key Values
42

Tables
0

Line Items
11

Table Extraction
No

Has been exported
No

Has been downloaded
No

Has been cleaned up
Yes

DeleteRelink AddressesSave

Invoices

Search: All Text ColumnsGoRechnungs-KenndatenActionsReset

Docu	File Name	Vendor Name	Vendor Address	Vendor Logo	Linked Vendor Address	Invoice Id	Subtotal	Total Vat	Invoice Total	Amount Due	Invoice Date	Total Tax	Due Date
	Oracle_Re_300002332.pdf	Oracle Austria GmbH	Donau-City-Strasse 7, WI...	ORACLE	Oracle Austria GmbH, Don...	300002332	1,050.00	210.00	1,260.00	1,260.00	22.11.2021		22.12.20
	Oracle_Re_300006785.pdf	0 Oracle Austria GmbH	Donau-City-Strasse 7, WI...	ORACLE	Oracle Austria GmbH, Don...	300006785	1,050.00	210.00	1,260.00	1,260.00	23.05.2022		22.06.20

1 rows selectedTotal 2

Invoice Line Items

Search: All Text ColumnsGoActions

Page

Line	Name	Quantity	Unit	Unitprice	Amount	Tax	Date	Product Code
Page: 1								
1	connectPM Plattformsteuerung von 01/11/21 bis 31/0...				1,050.00			
2				1,050.00	1,050.00			

Total 2

Export invoices

A click on 'Export to Accounting' displays a preview of the exported data. With a click on 'execute export', the displayed data is transferred to the accounting database as incoming invoices. The searchable Pdf files are transferred to the accounting database as attachments to the invoices. After the export, the button is hidden to prevent a double export.

Rechnungen in die Buchhaltungsdatenbank exportieren												
Job-ID: 000093												Export ausführen
Document_Id ↑	Job_Id	File_Name	Doc_Ustidnr	Workspace_Id	Eingangsdatum	Faelligkeitsdatum	Belegdatum	Belegjahr	Abrechnungsjahr	Posten_Nr	Forderung	Gaenderte_Forderung
242	93	230309_Octopus.pdf		124	05.06.2023	23.03.2023	05.04.2023	2023	2023	20230258	525,02	525,02
243	93	230329_Ing_Moser_0431_2023.pdf		124	05.06.2023		12.04.2023	2023	2023		8951,26	8951,26
244	93	230323_Archt_Werkstatt_RE_234_ZL_10.pdf	0043 5355 65898	124	05.06.2023		23.03.2023	2023	2023	234	3120	3120
245	93	230331_MBLD_RE_5522.pdf	583 0464 38	124	05.06.2023	31.03.2023	31.03.2023	2023	2023	5522	10990	10990
246	93	230401_Praschberger_RE_1010676.pdf		124	05.06.2023		01.04.2023	2023	2023		165	165
Herunterladen Drucken												

Document analysis

On the Document Results page, the searchable PDF view of the uploaded document is displayed on the left. The document can be downloaded as a PDF file in the lower area of the view.

At the top right you can choose from the possible evaluations. Depending on the recognized document type, document attributes, field labels, line items, invoice items, receipt line items, tables and text lines are displayed here.

OCI Document Understanding AI

Search Configuration dstrack

Home \ Processor Job Details \

Document Results

Process Another Job Details

Oracle_Re_300002332.pdf • 1 of 2

Next

1 von 1

Automatischer Zoom

ORACLE

RECHNUNG

Gesamtbetrag
1.260,00

Fälligkeit
22-DEC-2021

Rechnungsnummer
300002332

Rechnungsdatum
22-NOV-2021

Zahlungsbedingungen
Nett Due in 30 Days

Vertragsnummer
1000018455

Kunden UST-ID-NR.
ATU74785958

Steuerpunkt Datum
22-NOV-2021

Rechnungsnummer
300002332

Referenzrechnungsnummer: 300002332

1. Rechnungsgeber: ITS
Bayer Parkhaus Fürte SA
Bayer AT 100 100 0001 5414 0200
SWIFT Code: GEBAA1700000

Netto Gesamtbetrag
1.050,00

MwSt
(19%)
210,00

Gesamtbetrag
1.260,00 EUR

Rechnungsempfänger
Kitzbüheler Alps Projekt GmbH
Sebastian Strle
Obere Gänsbachgasse 7
6370 Kitzbühel
Austria

Leistungsanbieter
Kitzbüheler Alps Projekt GmbH

Phz

Rückbetrag

Menge

Einzelpreis

Netto-Gesamt-Preis

1

correctPM Plattformsteuerung von 01/11/21 bis 31/01/22 (550€ /Monat)

1

1.050,00

1.050,00

Document Attributes

Job ID
000172

Document-ID
369

File Name
Oracle_Re_300002332.pdf

Mime Type
application/pdf

Language Code
German

Language Score
75,4%

Document Type
Others

Document Type Score
99,0%

Job Language Code
English

Job Document Type
Invoice

Page Count
1

Key Values
21

Line Items
5

Tables
0

Total Amount
1.260,00

Matching Customer Address
Kitzbüheler Alps Projekt GmbH, Obere Gänsbachgasse 7 6370 Kitzbühel (72500)

Matching Vendor Address
Oracle Austria GmbH, Donau-City-Strasse 7 1220 Wien (73502)

Delete Save

Field Labels

Actions Edit Save Reset

Page

Confirmed

Field Label ↑ 2

Label

Field Alias

Field Text

Value Type

Number

document attributes

In addition to the document identification data (job ID, file name, MIME type), the recognized language and the document type are also displayed here. The language and the document type can be changed if necessary so that the appropriate number and date formats can be used in the evaluation. For invoices and receipts, the number of pages, field values, line items and tables, total amount and associated addresses are displayed.

Document Attributes

Job ID

000127

File Name

ATU.pdf

Language Code

German

Document Type

Others

Job Language Code

German

Page Count

1

Line Items

37

Processorjob Message

Total Amount

50.20

Matching Customer Address

Dr. Peter Zdesar & Partner (60571)

Matching Vendor Address

Holz-Studio Oberacher GmbH & Co KG (60294)

Mime Type

application/pdf

Language Score

100.0%

Document Type Score

98.7%

Job Document Type

Invoice

Key Values

8

Tables

0

Save

field labels

The field labels found by the processor are listed for the output of the invoice and receipt data. The user can check and improve the data associations found by the processor.

1. Several options

If fields occur more than once, the user can select the desired entry by clicking in the 'Confirmed' column.

2. Data Conversion

In the Numerical Value and Date Value columns, the field text or field value is output as a converted number or date. If the data is not converting as expected, the language for a document can be set to apply the appropriate format masks and currency symbols. Several variants can be used for each language. For example, if the language German was

selected for the processor job and the processor finds the language English for a document, then 4 variants are used for the recognition and conversion.

Language	Territory	Date format	Number format	Currency Symbol	Currency Code
English	AMERICA	MM/DD/RRRR	999,999,999.99	\$	USD
English	UNITED KINGDOM	DD-MON-RRRR	999,999,999.99	£	GBP
German	GERMAN	DD.MM.RRRR	999.999.999,99	€	EUR
German	AUSTRIA	DD.MM.RRRR	999 999 999.99	€	EUR

3. Field Labels mapping

The field name can be reassigned in the User Label column.

If there are several options for a field, the desired entry can be selected by clicking in the Confirmed column.

Field Labels

Primary Report

Actions

Edit

Save

Reset

Page

Multiple choices

	Confirmed	Field Label ↑±2	Label Score	Field Text	User Label	Value Type	Number Value	Date Value
▼ Page: 1								
<div></div>	<div></div>	CustomerAddress	96	Ansbacher Str. 58 10777 Berlin		STRING		
<div></div>	<div></div>	InvoiceId	76	0306039 6		STRING		
<div></div>	<div></div>	InvoiceTotal	72	50,20		NUMBER	50.20	
<div></div>	<div></div>	SubTotal	40	50,19		NUMBER	50.19	
<div></div>	<div></div>	VendorAddress	97	Yorckstr. 57, 10965 Berlin, schoeneberg@atu.de ...		STRING		
<div></div>	<div></div>	VendorAddressRecipient	99	A.T.U Auto-Teile-Unger GmbH & Co. KG Berlin 12		STRING		
<div></div>	<div></div>	VendorName	99	A.T.U Auto-Teile-Unger GmbH & Co. KG Berlin 12		STRING		
<div></div>	<div></div>	VendorNameLogo	85	ATN io Meisterwerkslatt		STRING		

1 rows selected

Total 8

line item

The individual items found by the processor are listed for the recognized document types invoice and receipt. If the user recognizes invoice fields in the list, a suitable entry can be selected with a click in the column User field designation. These fields are then listed as entries in the Field Labels list.

For example, the invoice amount can then be marked if the processor was unable to do so.

Line Items					
<div><div><div><div></div><div>Search: All Text Columns</div><div>Go</div></div><div>Actions</div><div>Edit</div><div>Save</div></div><div>Reset</div></div>					
<div><div><div></div><div>Page</div><div></div></div></div>					
	Line	Field Label	Value Type	User Label	Field Text
Page: 1					
<input checked="" type="checkbox"/>	1	Description	STRING		RADWECHSEL MIT WUCHTEN STAHL
<input type="checkbox"/>	1	ProductCode	STRING		1452
<input type="checkbox"/>	1	Name	STRING		RADWECHSEL MIT WUCHTEN STAHL
<input type="checkbox"/>	1	Quantity	NUMBER		4
<input type="checkbox"/>	1	UnitPrice	NUMBER		9,30
<input type="checkbox"/>	1	Amount	NUMBER		37,20
<input type="checkbox"/>	2	Description	STRING		RADEINLAGERUNG DURCH KUNDEN
<input type="checkbox"/>	2	ProductCode	STRING		RAD21
<input type="checkbox"/>	2	Name	STRING		RADEINLAGERUNG DURCH KUNDEN
<input type="checkbox"/>	2	Quantity	NUMBER		1
<input type="checkbox"/>	2	UnitPrice	NUMBER		0,00
<input type="checkbox"/>	2	Amount	NUMBER		0,00
<input type="checkbox"/>	3	Description	STRING		RADNABEN/RADANLAGEFLAECHEN REINIGEN
<input type="checkbox"/>	3	ProductCode	STRING		EIN08
<input type="checkbox"/>	3	Name	STRING		RADNABEN/RADANLAGEFLAECHEN REINIGEN
<input type="checkbox"/>	2	Quantity	NUMBER		4
1 rows selected					Total 37

invoice items

For the invoice document type, the list of invoice items is formed from the individual items.

Recognized and converted numbers and date fields are formatted uniformly.

Invoice Line Items

Actions

Page

Line	Name	Quantity						
1	RADWECHSEL MIT WUCHTEN STAHL	4.00		9.30	37.20			1452
2	RADEINLAGERUNG DURCH KUNDEN	1.00		0.00	0.00			RAD21
3	RADNABEN/RADANLAGEFLAECHE REINIGEN	4.00		2.50	10.00			EINO8
4	WICHTIGE SICHERHEITSRELEVANTE PRFPUNKTE	1.00		0.00	0.00			81036
5	OBERFLCHENDESINFEKTION Schlssel)	1.00		2.99	2.99			
6	nfinanzielle		Nettobet...		42.17			
7	eren, wir b		MwSt.		8.02			
Total 7								

Tables

The tables recognized by the processor in the documents are listed here.

The first row of a table is highlighted in the report view. In the popup field 'Table' a table can be selected from a list of all tables in the document.

Tables								
Table								
Q Los Aktionen								
2 Page Number, Table Number Markieren: Header								
Page Number: 1, Table Number: 1								
Col001	Col002	Col003	Col004	Col005	Col006	Col007	Col008	Col009
Open installment invoices:	Net	VAT (%)	Gross					
AR0000400052 30.11.2021	538.516,52 EUR	107.703,30 EUR (20,00)	646.219,82 EUR					
AR0000400054 31.12.2021	470.382,33 EUR	94.076,47 EUR (20,00)	564.458,80 EUR					
AR0000400068 26.01.2022	294.227,71 EUR	58.845,54 EUR (20,00)	353.073,25 EUR					
AR0000400078 11.03.2022	560.599,79 EUR	112.119,96 EUR (20,00)	672.719,75 EUR					
AR0000400083 12.04.2022	119.415,33 EUR	23.883,07 EUR (20,00)	143.298,40 EUR					
AR0000400091 10.05.2022	435.059,89 EUR	87.011,98 EUR (20,00)	522.071,87 EUR					
AR0000400096 08.06.2022	883.534,44 EUR	176.706,89 EUR (20,00)	1.060.241,33 EUR					
AR0000400106 13.07.2022	1.712.645,12 EUR	342.529,02 EUR (20,00)	2.055.174,14 EUR					
AR0000400112 16.08.2022	1.061.572,85 EUR	212.314,57 EUR (20,00)	1.273.887,42 EUR					
AR0000400117 13.09.2022	871.518,62 EUR	174.303,72 EUR (20,00)	1.045.822,34 EUR					
AR0000400123 07.11.2022	808.435,59 EUR	161.687,12 EUR (20,00)	970.122,71 EUR					
AR0000400137 14.12.2022	1.014.136,32 EUR	202.827,26 EUR (20,00)	1.216.963,58 EUR					
AR0000400142 10.01.2023	1.244.838,09 EUR	248.967,62 EUR (20,00)	1.493.805,71 EUR					
AR0000400149 07.02.2023	946.216,34 EUR	189.243,27 EUR (20,00)	1.135.459,61 EUR					
Page Number: 1, Table Number: 2								
Col001	Col002	Col003	Col004	Col005	Col006	Col007	Col008	Col009
Headquarters: Bahnhofstrasse 29	Chairman of the Board: Hans Lindner Managing Directors:	Bank: HypoVereinsbank, IBAN DE39						
94424 Arnstorf	Heinrich Bichner, Helmut Lang, Patrick Ossiander, Martin Weber	Landshut 7432 0073 0027 9846 73 980/1173						
Germany	Registered Office: Arnstorf	HYVEDEMM433						
Phone +49 8723 20-0	HRB 12135	BIC Tax-No.						
	Commercial Register Local Court: Landshut							

Lines of text

The text lines recognized by the processor in the documents are listed here.

If a line of text is recognized as a date or number according to the formats of the document language, then the corresponding entries are displayed in the Value Type, Number Value and Date Value columns.

Textzeilen

Suchen: Alle Textspalten

Los

Aktionen

Zurücksetzen

Seitenzahl

Zeilennr. ↑±2	Textzeile	Value Type	Number Value	Date Value
Seitenzahl: 1				
1	Octopus	STRING		
2	Werbegestaltung & Montagebau GmbH	STRING		
3	Kitzbühler Alps Projekt GmbH	STRING		
4	ATU74785958	STRING		
5	Kontakt:	STRING		
6	Obere Gnsbachgasse 7	STRING		
7	Wasserfeld 1a	STRING		
8	6370 Kitzbhel	STRING		
9	6361 Hopfgarten im Brixental	STRING		
10	Mobil: +43 /676/9616336	STRING		
11	Mobil: +43 /676 / 5339782	STRING		
12	Datum:	STRING		
13	09.03.2023	DATE		09.03.2023
14	Rechnungs-Nr.:	STRING		
15	20230258	NUMBER	20.230.258,00	
16	Kunden-Nr.:	STRING		
17	2265	NUMBER	2.265,00	
18	Sachbearbeiter/-in:	STRING		
19	Mathias Gandler	STRING		
Gesamt 169				

Text preview

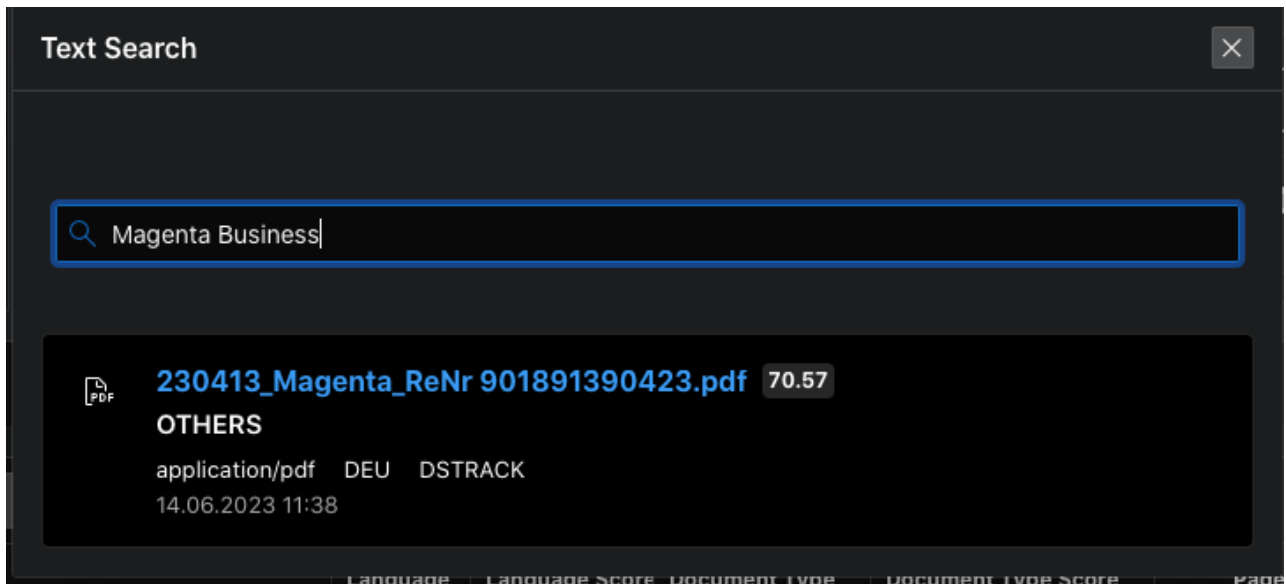
The entire text as used in the text index is output here.

Text Preview

Octopus Werbegestaltung & Montagebau GmbH Kitzbühler Alps Projekt GmbH ATU74785958 Kontakt: Obere Gnsbachgasse 7 Wasserfeld 1a 6370 Kitzbhel 6361 Hopfgarten im Brixental Mobil: +43 /676/9616336 Mobil: +43 /676 / 5339782 Datum: 09.03.2023 Rechnungs-Nr.: 20230258 Kunden-Nr.: 2265 Sachbearbeiter/-in: Mathias Gandler Rechnung Sehr geehrte Damen und Herren, wir erlauben uns, wie folgt in Rechnung zu stellen: Projekt: Sujetwechsel LED Rahmenit Anfrage vom 24.02.2023 Auftraggeber:Herr Walter Fischer Pos. Anzahl Einheit Artikelnr. Bezeichnung Rabatt Einzelpreis Gesamtpreis 1 1 Std. 099 Datenbernahme, Datenhandling inkl. RIP 50% 15,00 € 15,00 € 2 2,43 m2 134 hochwertiger Sublimations_Digitaldruck 25% 42,30 € 102,79 € auf Frametex B1 220 g Format : It Naturmanahme V2 Konfektion : rundum Gummilippe vernht 1 STK 3 6,3 lfm 212 Silikonflachkeder 10% 4,68 € 29,48 € 4 4,86 m2 134 hochwertiger Sublimations_Digitaldruck 25% 42,30 € 102,79 € auf Frametex B1 220 g Format : It Naturmanahme V2 Konfektion : rundum Gummilippe vernht 1 STK 5 99 lfm 212 Silikonflachkeder 10% 4,68 € 46,33 € 6 1 Stk. 900 Montage Sujetwechsels 30,00 € 30,00 € Montagepauschale incl. An- und Abfahrt zum Montageort Montagedatum: 03.02.2023 Montageort: Sujetwechsel Montageort : Kitzbhel vertrag 429,18 € octopus Werbegestaltung & Montagebau GmbH Wasserfeld 1a E-Mail: s.weinert@octopus-wm.at ; m.gandler@octopus-wm.at A-6361 Hopfgarten im Brixental UID-Nr. ATU 63536935 - FN 296280g Rechnungs-Nr.: 20230258 Seite 2 von 2 Pos. Anzahl Einheit Artikelnr. Bezeichnung Rabatt Einzelpreis Gesamtpreis bertrag 429,18 € Summe 429,18 € Mehrwertsteuer 20% auf 429,18 € netto 85,84 € Gesamtbetrag 515,02 € Zahlungsbedingung: Zahlbar nach Erhalt der Rechnung, ohne Abzug Es gilt die Zuständigkeit des Bezirksgerichts Kufstein und die AGB's octopus Werbegestaltung & Montagebau GmbH Die Ware bleibt bis zur vollständigen Bezahlung unser Eigentum. Bankverbindung: Sparkasse Wrgl-Kufstein, BLZ.: 20506, Konto Nr. 00100031731 IBAN: AT022050600100031731, BIC: SPKUAT22XXXX Unsere UID:ATU 63536935 ;FN 296280g EINGEGANGEN 0 6 April 2023 Octopus Werbegestaltung & Montagebau GmbH Bro Franz Wieser Kitzbhel Kitzbühler Alps Projekt GmbH ATU74785958 Kontakt: Obere Gnsbachgasse 7 Wasserfeld 1a 6370 Kitzbhel 6361 Hopfgarten im Brixental Mobil: +43 / 676 / 9616336 Mobil: +43 / 676 / 5339782 Datum: 05.04.2023 Sachbearbeiter/-in: Mathias Gandler Zahlungserinnerung Sehr geehrte Damen und Herren, leider mussten wir feststellen, dass folgende Rechnung Ihrerseits noch nicht beglichen wurde: 20230258 Rechnungs-Nr.: Rechnungsdatum: 09.03.2023 23.03.2023 Flig am: 515,02 € Rechnungsbetrag: Bisher gezahlt: 0,00 € 10,00 € Mahngebühren: 525,02 € Zu zahlender Betrag: Sollten Sie Ihrerseits die Rechnung schon beglichen haben,dann sehen sie dieses Schreiben als gegenstandslos an. Wir mchten Sie bitten, den noch ausstehenden Betrag umgehend zu berweisen. Es gilt die Zuständigkeit des Bezirksgerichts Kufstein und die AGB's octopus Werbegestaltung & Montagebau GmbH. Die Ware bleibt bis zur vollständigen Bezahlung unser Eigentum. Bankverbindung: Sparkasse Wrgl-Kufstein, BLZ.: 20506, Konto Nr. 00100031731 IBAN: AT022050600100031731, BIC: SPKUAT21 octopus Werbegestaltung & Montagebau GmbH Wasserfeld 1a E-Mail: s.weinert@octopus-wm.at ; m.gandler@octopus-wm.at A-6361 Hopfgarten im Brixental UID-Nr. ATU 63536935 - FN 296280g

Search documents

Clicking on the magnifying glass icon in the top right corner of the menu bar opens a dialog window for the full-text search. The searchable Pdf files stored in the object store can be quickly searched with an Oracle text index. By entering keywords, corresponding documents are listed. With a click in the list of search results, the document can be opened in the Document Analysis page.



Configuration

Clicking on the Configuration entry in the menu bar at the top right opens the dialog window for the AI processor configuration. In addition, the field types, line item types and languages are set up here as defined for the processor version in the manufacturer specifications and your Oracle Cloud console.

AI processor configuration

All necessary parameters for accessing the Oracle Object Store and the AI Document Understanding web service are set up here.

Edit AI Processor Configuration

Current Configuration

20221109/processorJobs

Show All

AI Processor Configuration

Field Types

Line Item Types

Languages

Text Index Status

Configuration Name

20221109/processorJobs

Input Location Prefix

scan_ai_documents

Output Location Prefix

scan_ai_results

Pdf Location Prefix

searchablePdf

Delete Processor Output

☒

Compartment Id

ocid1.compartment.oc1..aaaaaaaackpx37s32zo44j4ymjdfplq4uk6fnvkfl6qnj6azezaefqoefdfa

Object Bucket Name

APEX_OCI_AI_REST_SERVICE

Object Namespace Name

frsvel7ogr9o

Object Store Base Url

https://objectstorage.eu-frankfurt-1.oraclecloud.com

Oci Doc Ai Url

https://document.aiservice.eu-frankfurt-1.oci.oraclecloud.com/20221109/processorJobs

Wc Credential Id

APEX_OCI_CREDENTIAL

Currency Character

€

Number Character

., decimal point & group comma

Processor Language Code

English

Processor Documenttype

Invoice

Key Values Extraction

☒

Table Extraction

☐

Generate Searchable Pdf

☒

Keep Original Files

☐

Execute Asynchronous

☒

Context Id Query

SELECT NAME, ID FROM WECO_TOWER1.PROJEKTE_BT WHERE WORKSPACE\$_ID = 124 ORDER BY 1

Address Id Query

SELECT FIRMA1||' ('||MANDANTEND||')' NAME, ID FROM WECO_TOWER1.ADRESSEN_BT WHERE WORKSPACE\$_ID = 124 ORDER BY 1

Find Address Function

Document_Scan_AI_Find_Address

Invoice Export View

Weco_Tower1.V_DOCUMENT_SCAN_AI_INVOICE_EXPORT

Invoice Export Procedure

Weco_Tower1.EXPORT_INVOICE_DOCUMENTS

Cancel

Save a Copy

Apply Changes

Field-Name	Column-Name	Example Contents	Meaning
Configuration name		20221109/processorJobs	Name of the configuration

Field-Name	Column-Name	Example Contents	Meaning
Input location prefix	input_location_prefix	scan_ai_documents	Collection folder name for the uploaded files in the object storage
Output location prefix	output_location_prefix	scan_ai_results	Collection folder name for the AI processor results in the object storage
PDF location prefix	pdf_location_prefix	searchable_pdf	Collection folder name for searchable pdf documents in the PDF location prefix in the object storage
Compartment ID	compartment_id	ocid1.compartment.oc1..aaaaaaa	Compartment_Id in the object storage
Object bucket name	object_bucket_name	APEX_OCI_AI_REST_SERVI	Bucket name in the object storage
Object namespace name	object_namespace_name	CE frsvel7xgy9o	Object namespace name in object storage
Object storage base URL	object_store_base_url	https://objectstorage.eu-frankfurt-1.oraclecloud.com	Base_Url of the object store
OCI Doc AI Url	oci_doc_ai_url	https://document.aiservice.eu-frankfurt-1.oci.oraclecloud.com/20221109/processorJobs	Base_Url of the Document Understanding AI Processor
WC Credential ID	wc_credential_id	APEX_OCI_CREDENTIAL	Static name of the web credentials in the APEX workspace
Context ID Query	context_id_query	select name d, id r from projects	Value list - select statement for the assignment of a context for later processing
search function	find_address_function	Document_Scan_AI_Find_Addr	Function name for the address search
Address list query	address_id_query	ess Select name, id from addresses	Value List - select statement for the display of addresses
Invoice export view	invoice_export_view	Weco_Tower1.V_DOCUMENT_SCAN_AI_INVOICE_EXPORT	Name of the SQL view for Invoice export
Invoice export procedure	invoice_export_procedure	T Weco_Tower1.EXPORT_INVOICE_DOCUMENTS	Name of SQL procedure for invoice export
currency sign	currency_character	€	Währungszeichen Standardvorgabe
number character	number_character	,.	Default Decimal & group character
Processor language code	processor_language_code	English	Processor Languages Default. Must be English if key value extraction is enabled.
Processor document type	processor_documentType	Invoice	Document Type Default
key values extraction	Key_Values_Extraction	On	Processor Option Key Value Extraction - Generates field labels and line items for invoices, receipts, and field labels for passport, driver's license
table extraction	Table_Extraction	On	Option Table Extraction - Generates tables with information about the page, number of columns and rows from the document contents.
Generate searchable PDF	generate_Searchable_Pdf	On	OCI Processor Option Generate Searchable PDF - generate

Field-Name	Column-Name	Example Contents	Meaning
			searchable PDF from document content. The files are generated in the object store with the PDF location prefix. An Oracle Text Index makes the documents searchable with keywords.
Delete processor output	delete_processor_output	On	delete AI Processor results in object store after storing data in SQL table.
Keep original files	keep_original_files	Off	do not delete the uploaded files in the object store. when Generate Searchable PDF is turned on, the files can be searched even if the original files have been deleted.
Run asynchronously	execute_asynchronous	On	f is switched on during the execution of OCI Processor jobs, a dialog window with the processing progress and a forecast of the processing time in seconds is displayed. Also, this avoids a timeout error due to too long processing time.

Set up Oracle Text Index for the address table

In order to make the table with the field contents from the analysis result searchable, an Oracle Text Index must be set up for the address table. You can find an example of this in the

Accounting_Address_Text_Index.sql file.

Register the address lookup function

A function with the profile

FUNCTION Find_Address (p_Search IN VARCHAR2,p_Language IN VARCHAR2) RETURN NUMBER is installed in the application scheme and entered in the configuration (find_address_function). See the Accounting_Find_Address.sql file for an example.

Set up queries for context and addresses

Select queries for the address list (address_id_query) and project list (context_id_query) are entered.

Example for address_id_query: SELECT COMPANY1||' ('||TENANTID||')' NAME, ID FROM WECO_TOWER1.ADRESSEN_BT WHERE WORKSPACE\$_ID = 124 ORDER BY 1

Example of context_id_query: SELECT NAME, ID FROM WECO_TOWER1.PROJEKTE_BT WHERE WORKSPACE\$_ID = 124 ORDER BY 1

Set up export to accounting

In the Invoice export view field, the name of the SQL view for the export view and in Invoice export procedures the name of the SQL procedure for the invoice export to accounting is set up. The export view shows the data compilation resulting from the context (project), the invoice data and the addresses found. For a job ID, the procedure transfers the data from the export view to the accounting database as incoming invoices with attachments. You can find an example of this in the Accounting_Invoice_Export.sql file.

Manage multiple configurations

The current configuration can be created as a duplicate by clicking on 'Save a copy'.

Rename by changing the content of the Label field.

Custom settings for processor configuration, field types, line item types and languages can be made in the duplicate.

If you have multiple configurations, you can select one from the list at the top of the page and mark it as the current configuration.

Configuration - Field Types

All field names of the key value extraction for the individual document types are defined here. Their meaning, value type, and format masks as defined for the processor version in the manufacturer specification are set up here. For the assignment of the field types to a document, the document type of the order and the document type determined for each document are taken into account and valid field types are adopted in the results output.

Edit AI Processor Configuration

Current Configuration
20221109/processorJobs

Show AllAI Processor ConfigurationField TypesLine Item TypesLanguagesText Index Status

Field Types

Q

Search: All Text Columns

Go

Actions

Edit

Save

Add Row

Reset

Document Type

	Field Label ↑ ±2	Description	Api Response Value	Value Type	Format Mask
▼	Document Type: DRIVER_LICENSE				
<input checked="" type="checkbox"/>	Address	Address listed on the document.	Extracted Text	STRING	
<input type="checkbox"/>	BirthDate	Date of birth.	Date in YYYY/MM/DD format	DATE	YYYY/MM/DD
<input type="checkbox"/>	Country	Country listed on the document.	Extracted ISO 3166-1 country code	STRING	
<input type="checkbox"/>	DocumentNumber	Document identification number.	Extracted Text	STRING	
<input type="checkbox"/>	ExpiryDate	Date of expiration listed on the docu...	Date in YYYY/MM/DD format	DATE	YYYY/MM/DD
<input type="checkbox"/>	FirstName	First name (given name) listed on th...	Extracted Text	STRING	
<input type="checkbox"/>	Gender	Gender listed on the document.	Extracted Text	STRING	
<input type="checkbox"/>	IssueDate	Date of issue listed on the document.	Date in YYYY/MM/DD format	DATE	YYYY/MM/DD
<input type="checkbox"/>	LastName	Last name (family name) listed on th...	Extracted Text	STRING	
<input type="checkbox"/>	Region	Region listed on the document. For ...	Extracted Text	STRING	
▼	Document Type: INVOICE				
<input type="checkbox"/>	AmountDue	Total amount due to the vendor.		NUMBER	
<input type="checkbox"/>	BillingAddress	Explicit billing address for the custo...		STRING	
<input type="checkbox"/>	BillingAddressRecipient	Name referenced with the BillingAd...		STRING	
<input type="checkbox"/>	CustomerAddress	Mailing address for the Customer.		STRING	
<input type="checkbox"/>	CustomerAddressRecipient	Name referenced with the Customer...		STRING	
<input type="checkbox"/>	CustomerId	Customer reference identifier.		STRING	
<input type="checkbox"/>	CustomerName	Name of invoiced customer.		STRING	
<input type="checkbox"/>	DueDate	Date when payment is due on this in...		DATE	

1 rows selectedTotal 60

Cancel

Save a Copy

Apply Changes

Configuration - Line Item Types

All field names of the line item extraction for the document types invoice and receipt are defined here. Its meaning and value type as defined for the processor version in the manufacturer specification are set up here.

Edit AI Processor Configuration

Current Configuration
20221109/processorJobs

Show AllAI Processor ConfigurationField TypesLine Item TypesLanguagesText Index Status

Line Item Types

Search: All Text ColumnsGoActionsEditSaveAdd RowReset

Document Type

<input type="checkbox"/>	<input type="checkbox"/>	Line Item ↑±2	Description	Value Type
Document Type: INVOICE				
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Amount	The amount of the line item.	NUMBER
<input type="checkbox"/>	<input type="checkbox"/>	Date	Date corresponding to each line item.	DATE
<input type="checkbox"/>	<input type="checkbox"/>	Description	The text description for the invoice line item, for example, men's ...	STRING
<input type="checkbox"/>	<input type="checkbox"/>	Items	Concatenation of all other line item values (that is, the entire line ...	STRING
<input type="checkbox"/>	<input type="checkbox"/>	Name	The name listed for a product or service, for example, t-shirt.	STRING
<input type="checkbox"/>	<input type="checkbox"/>	ProductCode	Product code, product number, or SKU referenced in the line ite...	STRING
<input type="checkbox"/>	<input type="checkbox"/>	Quantity	The quantity for this invoice line item.	NUMBER
<input type="checkbox"/>	<input type="checkbox"/>	Tax	Tax explicitly referenced on the line item.	NUMBER
<input type="checkbox"/>	<input type="checkbox"/>	Unit	The unit of the line item, for example, kg, lb.	STRING
<input type="checkbox"/>	<input type="checkbox"/>	UnitPrice	The price per item identified on the invoice.	NUMBER
Document Type: RECEIPT				
<input type="checkbox"/>	<input type="checkbox"/>	ItemName	Name of the item.	STRING
<input type="checkbox"/>	<input type="checkbox"/>	ItemPrice	Unit price of the item.	NUMBER
<input type="checkbox"/>	<input type="checkbox"/>	ItemQuantity	The number of each item purchased.	NUMBER
<input type="checkbox"/>	<input type="checkbox"/>	ItemTotalPrice	The total price of the line item.	NUMBER

1 rows selectedTotal 14

CancelSave a CopyApply Changes

Configuration - Languages

The list of languages includes NLS parameters with which the texts in the processor output are converted to Oracle Number and Date data types if the data matches the parameters. Several lines can be defined for each ISO code. In order to convert the correct data from the field contents of the number and date type, the language of the order and the language determined for each document are used to determine the conversion parameters. If, for example, the language of the order is German and the determined language of a document is English, then the lines with the areas AMERICA, UNITED KINGDOM and GERMANY are considered for the conversion and valid conversions are included in the output results.

Edit AI Processor Configuration

Current Configuration
20221109/processorJobs

Show AllAI Processor ConfigurationField TypesLine Item TypesLanguagesText Index Status

Languages

Q

Search: All Text Columns

Go

Actions

Edit

Save

Add Row

Reset

		ISO Code	Language Name	Nls Date Language	Common Date Form	Common Number Fo	Territory	Nls Numeric Charac	Nls Currency	Nls Iso Currency
<input checked="" type="checkbox"/>		BGR	Bulgarian	BULGARIAN	DD.MM.RRRR	999 999 999,99	BULGARIA	, decimal comma ...	лв	BGN
<input type="checkbox"/>		CHS	Simplified Chinese	SIMPLIFIED CHIN...	RRRR-MM-DD	999,999,999.99	CHINA	., decimal point &...	¥	CNY
<input type="checkbox"/>		CHT	Traditional Chinese	TRADITIONAL CH...	RRRR-MM-DD	999,999,999.99	TAIWAN	., decimal point &...	NT\$	TWD
<input type="checkbox"/>		CZE	Czech	CZECH	DD.MM.RRRR	999 999 999,99	CZECH REPUBLIC	, decimal comma ...	Kč	CZK
<input type="checkbox"/>		DEU	German	GERMAN	DD.MM.RRRR	999.999.999,99	GERMANY	., decimal comma...	€	EUR
<input type="checkbox"/>		DEU	German	GERMAN	DD.MM.RRRR	999 999 999,99	AUSTRIA	., decimal point & ...	€	EUR
<input type="checkbox"/>		DNK	Danish	DANISH	DD-MM-RRRR	999.999.999,99	DENMARK	., decimal comma...	kr	DKK
<input type="checkbox"/>		ENG	English	ENGLISH	DD-MON-RRRR	999,999,999.99	UNITED KINGDOM	., decimal point &...	£	GBP
<input type="checkbox"/>		ENG	English	AMERICAN	MM/DD/RRRR	999,999,999.99	AMERICA	., decimal point &...	\$	USD
<input type="checkbox"/>		ESP	Spanish	SPANISH	DD/MM/RRRR	999.999.999,99	SPAIN	., decimal comma...	€	EUR
<input type="checkbox"/>		FIN	Finnish	FINNISH	DD.MM.RRRR	999 999 999,99	FINLAND	, decimal comma ...	€	EUR
<input type="checkbox"/>		FRA	French	FRENCH	DD/MM/RRRR	999 999 999,99	FRANCE	, decimal comma ...	€	EUR
<input type="checkbox"/>		GRC	Greek	GREEK	DD/MM/RRRR	999.999.999,99	GREECE	., decimal comma...	€	EUR
<input type="checkbox"/>		HRV	Croatian	CROATIAN	DD.MM.RRRR	999.999.999,99	CROATIA	., decimal comma...	kn	HRK
<input type="checkbox"/>		HUN	Hungarian	HUNGARIAN	RRRR.MM.DD	999 999 999,99	HUNGARY	, decimal comma ...	Ft	HUF
<input type="checkbox"/>		ITA	Italian	ITALIAN	DD/MM/RRRR	999.999.999,99	ITALY	., decimal comma...	€	EUR
<input type="checkbox"/>		JPN	Japanese	JAPANESE	RRRR/MM/DD	999,999,999.99	JAPAN	., decimal point &...	¥	JPY
<input type="checkbox"/>		KOR	Korean	KOREAN	RRRR-MM-DD	999,999,999.99	SOUTH KOREA	., decimal point &...	₩	KRW
<input type="checkbox"/>		NLD	Dutch	DUTCH	DD-MM-RRRR	999.999.999,99	THE NETHERLAN...	., decimal comma...	€	EUR
<input type="checkbox"/>		NOR	Norwegian	NORWEGIAN	DD.MM.RRRR	999 999 999,99	NORWAY	, decimal comma ...	kr	NOK
<input type="checkbox"/>		POL	Polish	POLISH	RRRR-MM-DD	999 999 999,99	POLAND	, decimal comma ...	zł	PLN

1 rows selectedTotal 29

Cancel

Save a Copy

Apply Changes

Configuration - Text Index Status

The status of the text index is listed here for all documents. With a click on 'Rebuild Text Index' the index can be rebuilt if necessary.

Edit AI Processor Configuration

Current Configuration
20221109/processorJobs

Show All AI Processor ConfigurationField TypesLine Item TypesLanguagesText Index Status

Text Index Status

Q

Go

Actions

Rebuild Text Index

Document Id	Job Id	File Name	Has Text Index	Text Index Err Text
270	120	USCIS_US-Passport-Inside_180x241@2x.jpg	Y	
236	92	230413_Magenta_ReNr 901891390423.pdf	Y	
181	77	2023_05rechnung_5611287364.pdf	Y	
286	127	Booking .pdf	N	
182	77	Deine Rechnung von Apple für iCloud.pdf	Y	
238	92	230416_Salzburg_AG_RE_733715583.pdf	Y	
180	76	2023_04rechnung_5611287364.pdf	Y	
183	77	Deine Rechnung von Apple für Angry Birds.pdf	Y	
277	123	HASPA Kontoauszüge .pdf	N	
239	92	230418_Hofmann_RE_20573.pdf	Y	
195	81	Starbucks.jpeg	Y	
196	81	Walgreens.jpg	Y	
237	92	230414_Oesterreichische_Bundesforste_500228003.pdf	Y	
200	85	Chipotle.jpeg	Y	
284	127	Reifen Discounter .pdf	N	
260	115	Chipotle.jpeg	Y	
245	93	230331_MBLD_RE_5522.pdf	Y	
283	126	Quittung nachos.pdf	N	
266	119	Requirements-to-get-driving-license-for-H4-Visa-holders-No-SSN.jpg	Y	
240	92	230420_WF_RE_CD218B8A_0003.pdf	Y	
285	127	ATU.pdf	N	
213	89	Rechnung_2023_04_30254912000791.pdf	Y	

Cancel

Save a Copy

Apply Changes

Installation

Requirements

Prerequisites are a paid Oracle Cloud environment with an admin account to set up the required services and permissions.

The following schema permissions are required to install the supporting objects:

As SYS or ADMIN User:

GRANT EXECUTE ON CTXSYS.CTX_DDL TO <APP_SCHEMA>; – for Oracle Text Index

GRANT EXECUTE ON SYS.DBMS_LOCK TO <APP_SCHEMA>; – For Wait in asynchronous execution

GRANT EXECUTE ON DBMS_CLOUD TO <APP_SCHEMA>; – For Object Store access

OCI configuration

Security_Group, IAM User, API keys, Compartments, Policies and Object Storage Bucket must be configured in the OCI Service Console. Web credentials must be configured in APEX.

Example OCI configuration

Identity / Domains / Default domain /

Groups

Name: APEX_OCI_AI_REST_Security_Group

Description: Security Group for access to AI Services

Users

User Type: IAM User

Capabilities:

API keys: Yes

Auth tokens: Yes

Name: APEX_OCI_AI_REST_Service_Account

Description: Application User for AI REST Services

Email: your.name@example.de

Groups: APEX_OCI_AI_REST_Security_Group

API keys:

Download Private Key, Download Public Key

Configuration file - Copy

[DEFAULT]

user=ocid1.user.oc1..aaaa...

fingerprint=66:40:...

tenancy=ocid1.tenancy.oc1..aaaa...

region=eu-frankfurt-1

key_file=<path to your private keyfile> # TODO

Auth tokens:

Description : Access Buckets – Authentication token for
DBMS_CLOUD.CREATE_CREDENTIAL

Generate Token - Copy:

;...-b<FvGe

username => 'APEX_OCI_AI_REST_Service_Account'

password => ';...-b<FvGe'

Identity / Compartments

Name : APEX_OCI_SERVICES

Description: Compartment for Secure APEX Access to OCI Object Storage and AI REST Service

OCID <Copy> : ocid1.tenancy.oc1..aaa...

Identity / Policies

Name: APEX_OCI_AI_REST_Bucket_Policy

Description: Policy to assign APEX Access to OCI Services

Compartment: strackdev (root)/APEX_OCI_SERVICES

Policy Builder - Show manual editor:

Statements:

allow group APEX_OCI_AI_REST_Security_Group to read buckets in compartment APEX_OCI_SERVICES

allow group APEX_OCI_AI_REST_Security_Group to manage objects in compartment APEX_OCI_SERVICES where all {target.bucket.name='APEX_OCI_AI_REST_SERVICE', any {request.permission='OBJECT_CREATE', request.permission='OBJECT_INSPECT', request.permission='OBJECT_READ', request.permission='OBJECT_OVERWRITE'}}

allow group APEX_OCI_AI_REST_Security_Group to manage object-family in compartment APEX_OCI_SERVICES

allow group APEX_OCI_AI_REST_Security_Group to manage ai-service-document-family in compartment APEX_OCI_SERVICES

Object Storage

Compartment: APEX_OCI_SERVICES

Buckets: APEX_OCI_AI_REST_SERVICE

Default Storage Tier: Standard

Visibility: Private

Namespace <Copy>: frmnolrlgf8w

APEX Builder / Workspace Utilities / Web Credentials

Name: APEX OCI Credential

Credential Static ID: APEX_OCI_CREDENTIAL

Authentication Type: Oracle Cloud Infrastructure (OCI)

OCI User ID <API Keys Configuration file user>: ocid1.user.oc1..aaa...

OCI Private Key <API Keys private key file CR removed>: MIIEvQIBAD...

OCI Tenancy ID <API Keys Configuration file tenancy>: ocid1.tenancy.oc1..aaaa...

OCI Public Key Fingerprint <API Keys Configuration file fingerprint>: 5e:7b:...

Document Understanding API

Lookup your URL: <https://docs.oracle.com/en-us/iaas/api/#/en/document-understanding/20221109/>

API Endpoint: <https://document.aiservice.eu-frankfurt-1.oci.oraclecloud.com/20221109>

Install APEX application

Import the Oracle APEX installation file OCI_Document_Understanding_App.sql

Install the application and the supporting objects.

After installation as APP_SCHEMA user:

```
GRANT SELECT ON V_DOCUMENT_SCAN_AI_INVOICES TO <ACCOUNTING_SCHEMA> WITH GRANT OPTION;
```

```
GRANT SELECT ON V_DOCUMENT_SCAN_AI_LINE_ITEM_INVOICE TO <ACCOUNTING_SCHEMA> WITH GRANT OPTION;
```

```
GRANT SELECT ON V_DOCUMENT_SCAN_AI_DOC_FILES TO <ACCOUNTING_SCHEMA> WITH GRANT OPTION;
```

After installation as ACCOUNTING_SCHEMA user:

```
GRANT SELECT ON V_DOCUMENT_SCAN_AI_INVOICE_EXPORT TO <APP_SCHEMA>;
```

```
GRANT EXECUTE ON Export_Invoice_Documents TO <APP_SCHEMA>;
```

Processing of processor orders

1. Store files in the object store

The files uploaded in the job are stored in a subdirectory with a job number according to the configuration bucket name and input location prefix (input_location_prefix). The original file names are used after a conversion of special, accent and space characters. The conversion is necessary because the permitted character set is restricted.

2. Compile and start the processor job

The list of uploaded file locations and the specifications like document type, language and recognition features like text extraction, table extraction, document classification, key value extraction are sent to the processor as a job.

3. Awaiting processing

If the configuration **Run asynchronously** is activated, a scheduler job is started that updates the processor job status. The processor job status is called up in the user interface every 2 seconds and a dialog window informs about the status and the completed percentage of the job during execution. When the job is done, processing continues.

4. Processing of the results

After the successful completion of a processor job, the recognized language and document type, field labels with field contents, line items and tables are extracted from the analysis result of the processor job. Numbers and dates are converted according to country-specific formats because the AI processor only knows American formats.

The country-specific date formats, currency symbols, currency code, and decimal and group sign of the languages of the processor order and the found language are used. Also special date formats used for field labels and line items in the processor results.

When validating the conversion, extended conditions are checked to prevent, for example, €51,234.50 not being recognized as 51.2345 and 04/11/23 not being recognized as November 4th, 2023.

For numbers, the decimal point must occur only once and after the group characters, and the group character must occur at least 3 characters before the end and the distance to the decimal point must be a multiple of 3. Appropriate currency symbols, currency code, and the group sign may optionally be included. A dynamic format mask is therefore used for the number conversion.

The path to uploaded files or searchable pdf files is stored in the database in such a way that an Oracle Text Index can be created for the files. Because the accented characters and umlauts are missing in the recognized texts, the documents must be searched through as a vague search.

The field contents are condensed, converted and summarized and then stored in database tables. These are then available as lists for download.

5. Algorithm for Analyzing Text Fragments

The Document Understanding AI provides as a result of a Document Scan in a JSON file, in addition to the key-value pairs (for form fields with content), also text fragments with X/Y coordinates for text lines and table cells.

The key-value results of the Document AI are not sufficient for import into accounting. Therefore, an algorithm was developed that improves the key-value results.

The algorithm uses alias names, data type validation, and relative positions to form valid key-value pairs.

1. Problem: The key-value pairs for invoices contained in the analysis result have too few fields to make a unique assignment to clients/invoice issuers. For the recognition of the invoice issuer, only the fields VendorName, VendorAddress, and VendorTaxId are available. A full-text search with Fuzzy Search does not provide usable address assignments to an existing address list.

1.1. Extension of the field list:

Additional fields such as **VendorEmail**, **VendorPhone**, **VendorMobile**, **BankBIC**, and **BankIBAN** were added for address assignment.

1.2. Addition of addresses:

The addresses of the project participants were supplemented so that address, VAT ID number, and IBAN as well as SWIFT-BIC are present in a bank connection.

1.3. The search for the addresses of the invoice issuers now includes the additional fields, which enables a unique and reliable assignment.

2. Problem: The found key-value pairs for invoices are incomplete or incorrectly assigned.

2.1. Filtering based on context values:

If a context (project) has been defined for the process, the field contents such as VendorName, VendorEmail, VendorPhone, VendorTaxId, and BankIBAN can be compared with the values of the context (Client_Name, Client_Email, Client_Phone, Client_Tax_Id, Client_IBAN). If matches are found, assignments to the Vendor fields can be recognized as incorrect and excluded.

2.2. Improved search for field labels:

A new table DOCUMENT_SCAN_AI_FIELD_ALIAS with alias names for field labels of the key-value pairs enables a more precise and comprehensive field assignment. Specific alias names have been set up for German and English invoices.

For German invoices, the following field alias names have been set up:

AmountDue: Zahlungsbetrag, Zu zahlender Betrag, Zu zahlender Betrag in Euro, berweisungsbetrag
BankBIC: BIC, BIC (SWIFT-Code) der Empfängerbank, BIC(SWIFT-Code) der Empfängerbank, BIC(SWIFT-Code) der Empfängerbank, BIC-Code, BIC., SWIFT, SWIFT Code, SWIFT/BIC, Swift, Swift Code
BankIBAN: IBAN, IBAN (EUR), IBAN Code, IBAN Nummer, IBAN code, IBANEmpfängerin
CustomerId: Ihre Kundennummer, Kunden-Nr, Kunden-Nr., Kundennr, Kundennr., Kundennummer
DueDate: Fällig am, Fälligkeit, Fälligkeitsdatum, Zahlungstermin
InvoiceDate: Datum, Leistungsdatum, Rechnungsdatum
InvoiceId: Belegnummer, Re.- Nr., Rechn.-Nr., Rechnung, Rechnung Nr., Rechnungs-Nr., Rechnungsnr., Rechnungsnummer, Zahlungsreferenz
InvoicePaidDate: BEZAHLT
InvoiceReceiptDate: EINGANG, EINGEGANGEN, Eingang
InvoiceTotal: Endbetrag EUR (brutto), Gesamt, Gesamtbetrag, Gesamtbetrag (inkl. USt), Gesamtbetrag brutto, Gesamtsumme, Rechnungsbetrag, Rechnungsbetrag in EUR, Summe, Summe Brutto (EUR)
PaymentTerm: Zahlungsbedingung, Zahlungsbedingungen, Zahlungsziel
SubTotal: Betrag Netto, Betrag netto, Netto, Netto Gesamtbetrag, Nettobetrag, Nettosumme, Rechnungsbetrag netto
TotalTax: Summe USt (EUR)
TotalVAT: 20 % MWST, 20 % USt, Betrag Ust. 20%, MwSt, MwSt.
VendorEmail: E-MAIL, E-Mail, E-Mail-Adresse, E-mail, Email
VendorMobil: Mobil
VendorName: Abs, Absender, Account Owner, Empfängerin Name/firma, EmpfängerIn, EmpfängerIn Name/Firma, Name des Empfängers
VendorPhone: TEL., Tel., Telefon
VendorTaxId: UID, UID Nr, UID Nr., UID-Nr., UID-Nummer, UST-ID NR., USt.-ID, Umsatzsteuer-ID, Unsere UID, Unsere UID- Nummer, Unsere UID-Nummer

For English invoices, the following field alias names have been set up:

AmountDue: Amount due, Amount payable, Total Amount due, Total amount due
BankBIC: BIC, Swift Code
BankIBAN: IBAN, IBAN code
CustomerId: Customer number
DueDate: Date due, Due date, Payment Due
InvoiceDate: Date of issue, Invoicing date
InvoiceId: Invoice Number, Invoice number
InvoiceTotal: TOTAL, Total, Total amount
SubTotal: Subtotal
TotalVAT: 20% VAT
VendorEmail: E-mail, Email

VendorName: Account Name

VendorPhone: Phone

VendorTaxId: Our VAT ID, Tax number

In the results of the AI scan, text fragments with precise coordinates for recognized texts are present. By finding alias names and searching for text fragments that are located relatively to the right or below, key-value pairs are formed.

With pattern recognition, alias names and field values are determined:

- If a value in a text fragment follows an alias name, the field value is extracted to the right of the colon.
- Composite text fragments that contain several alias names and values are separated at a semicolon, vertical bar, period, or comma, and the names and values are extracted.
- If an alias name appears in a text fragment and in a second text fragment to the right of the alias name in the same line, the last value (for numbers) or the next value (otherwise) is used.
- If an alias name appears in a text fragment and in a second text fragment below the first in the same column, left- or right-aligned, extraction takes place.
- The potential field contents are checked for their match with the data types (EMAIL, IBAN, SWIFTBIC, PHONE, INITCAP, NUMERIC, ALPHANUM), and unsuitable data are filtered out.

3. Problem: The found key-value pairs for NUMBER and DATE are not correctly converted according to the language and region.

3.1 Based on the preset language for the scan job and the language determined by the Document AI, valid formats for date and number values are determined and validated. If multiple valid hits for a field are found, the largest value is preferred.

4. Addition of the Key-Value Pairs

The algorithm was implemented as an Oracle View. This view provides possible hits for the search for key-value pairs with valid alias names, data types, and validations for amounts and dates per import job and document. The best hits per field are selected from the possible hits and added to the key-value list. Based on this key-value list, the data lines for export to accounting are created.

5. Continuous Improvement of Results

5.1 Before exporting the invoice data lines, a list of new alias names can be checked. This makes sense especially when the invoice scans from a certain invoice issuer are imported for the first time and the displayed invoice data appear incomplete. New field names ending with a colon are automatically found. In addition, possible field names are listed that stand before or above a valid date or number value.

Click on the 'Assign Key Aliases' button on the 'Details of the Processor Order' page. If designations are found in the displayed list that can be assigned to a known field type, then this assignment can

be carried out with a click on the 'Assign Aliases' button for selected lines and the selected field type. Uninteresting designations can be hidden by choosing the field type *Ignore.

5.2 After performing an import job with PDF documents, the assigned personal account is displayed in the 'Address of the linked provider' column in the Invoices list. Check the assignments. Note that the assignment of a personal account in the data collection works best if the address data of the personal accounts also included email, phone, VAT ID number, IBAN, and Swift-BIC.

If you have supplemented the address data or alias name assignment and want to perform a reassignment, click on 'Re-link Addresses' on the 'Details of the Processor Order' page. After the reassignment, the data lines for the list of invoices are updated to include the new fields.

5. Cleanup of processor job results

The output files of a processor job are deleted after the contents have been saved to the database.

The searchable pdf files generated by the processor job are stored in the directory configured for this (pdf_location_prefix) in a subdirectory with the job number.

The original files stored in the object store are deleted if configured to do so and a searchable pdf file was generated by the processor job.

6. Make connections to the address table

The field contents for recognized invoices and receipts are used to automatically find matching entries in an address list and assign them to the document attributes. The procedure Document_Scan_Ai_Pkg.Link_Document_Addresses searches for matching entries in an address list and enters them in the document attributes CustomerAddress_ID, VendorAddress_ID, BillingAddress_ID, ShippingAddress_ID, ServiceAddress_ID and RemittanceAddress_ID. Because the accented characters and umlauts are missing in the recognized texts, the search with the field contents for suitable addresses must be carried out in a fuzzy manner (FUZZY).