



Bank Statement Labeling Instructions

Evaluation Set Requirements



In order to do fair and thorough evaluation of the model, the evaluation set should take the following dimensions into consideration

1. **Revisions:** Documents should be in English across all the variations.
2. **Size:** It is recommended to have 500+ documents for each evaluation set.
3. **Template Diversity:** The documents should be diverse, e.g. different templates, issue year, different sources ... etc.
4. **Value Diversity:** Documents should capture all possible values.
5. **Handwritten:** It is preferred to separate handwritten documents from typed ones (i.e. have 2 separate evaluation sets) and report metrics for each.

Processor Version

These instructions are specific for Bank Statement Processor version ID:
“pretrained-bankstatement-v3.0-2022-05-16”

bank-statement [DISABLE PROCESSOR](#) [ACTIVITY](#)

OVERVIEW DATASET EVALUATE **MANAGE VERSIONS** HUMAN-IN-THE-LOOP

Default version
Google Release Candidate

The default version is used to process documents posted to your processor's prediction endpoint URL.

Versions [DEPLOY](#) [UNDEPLOY](#) [COMPARE](#) [REFRESH TABLE](#)

Filter Filter versions

<input type="checkbox"/>	<input checked="" type="radio"/>	Version ID	Created ↓	Status	Name	F1 score ?	API
<input type="checkbox"/>	<input checked="" type="radio"/>	pretrained-2022-05-16	Jul 30, 2022, 5:21:03 AM	Deployed	Google Release Candidate	RUN EVALUATION FIRST	VIEW DETAILS SAMPLE REQUEST ⋮
<input type="checkbox"/>	<input checked="" type="radio"/>	pretrained-next	Jul 30, 2022, 5:21:03 AM	Deployed	Google Default Next	RUN EVALUATION FIRST	VIEW DETAILS SAMPLE REQUEST ⋮
<input type="checkbox"/>	<input checked="" type="radio"/>	pretrained-next-v2	Jul 30, 2022, 5:21:03 AM	Deployed	Google Default Next V2	RUN EVALUATION FIRST	VIEW DETAILS SAMPLE REQUEST ⋮
<input type="checkbox"/>	<input checked="" type="radio"/>	pretrained	Jul 30, 2022, 5:21:03 AM	Deployed	Google default	RUN EVALUATION FIRST	VIEW DETAILS SAMPLE REQUEST ⋮

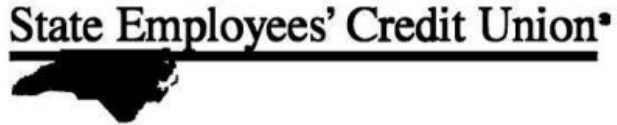
Entity List

Field	Description	Occurrence type	Base Type
account_number	Account Number.	OPTIONAL_MULTIPLE	string
account_type	Type of Account.	OPTIONAL_MULTIPLE	string
bank_address	Address of the bank.	OPTIONAL_ONCE	address
bank_name	Name of the bank.	OPTIONAL_ONCE	string
client_address	Address of the client.	OPTIONAL_ONCE	address
client_name	Name of the client.	OPTIONAL_MULTIPLE	string
ending_balance	The ending balance is the net residual balance in an account.	OPTIONAL_MULTIPLE	money
starting_balance	A starting balance is the amount of funds in an account at the beginning of a new fiscal period.	OPTIONAL_MULTIPLE	money
statement_date	Statement date is the date on which the statement is generated every month.	OPTIONAL_ONCE	datetime
statement_end_date	Statement ending date is typically the last day of your monthly billing cycle.	OPTIONAL_ONCE	datetime
statement_start_date	Date specified in the "Statement of Account" on which the statement is generated and printed.	OPTIONAL_ONCE	datetime
table_item	One row of a table encompassing the other table item values for table_item/*entities.	OPTIONAL_MULTIPLE	string
table_item/transaction_deposit	Transaction deposit amount.	OPTIONAL_ONCE(At most once per table item)	money
table_item/transaction_deposit_date	Transaction deposit date.	OPTIONAL_ONCE(At most once per table item)	datetime
table_item/transaction_deposit_description	Transaction deposit description.	OPTIONAL_ONCE(At most once per table item)	string
table_item/transaction_withdrawal	Transaction withdrawal amount.	OPTIONAL_ONCE(At most once per table item)	money
table_item/transaction_withdrawal_date	Transaction withdrawal date.	OPTIONAL_ONCE(At most once per table item)	datetime
table_item/transaction_withdrawal_description	Transaction withdrawal description.	OPTIONAL_ONCE(At most once per table item)	string

statement_date

Note : If only one date is present, then annotate it as "statement_date". In that case "statement_start_date" and "statement_end_date" is not present.

01



statement_date
Statement Date: 02-20-19
Page 1

Alexander Snyder
Mr. Joshua Wallace
57 Campos Course
Livingston NJ, 07039

Beginning Balance/Date - Example

Transaction Date	Effective Date	Description of Transactions	Transaction Amount	Balance
03/01		ID 01 MEMBERSHIP SAVINGS <u>Beginning Balance</u>		1707.50
03/14		Withdrawal Home Banking Transfer To NNNN ,NNNN C XXXXXXXXXXXX Share 12 Funds Transfer via Online	40.00	1667.50
03/26		Deposit Home Banking Transfer From Share 14 Funds Transfer via Online	100.00	1767.50
03/26		Deposit Mobile Deposit	78.60	1846.10
03/31		Deposit Dividend DIVIDEND	0.30	1846.40
		Annual Percentage Yield Earned 0.21% from 03/01/20 through 03/31/20 Based on Average Daily Balance of 1,718.84		
03/31		Ending Balance		1846.40
		Dividends Paid Year to Date	0.56	
		Dividends Paid In 2019	0.08	
Withdrawals and Other Charges				
Date	Amount	Description	Date	Amount Description
03/14	40.00	Withdrawal Home Banking		
Deposits and Other Credits				
Date	Amount	Description	Date	Amount Description
03/26	100.00	Deposit Home Banking Tra	03/31	0.30 Deposit Dividend
03/26	78.60	Deposit		
3 Deposits or Other Credits for 178.90				
03/01		ID 14 VALUE CHECKING <u>Beginning Balance</u>		880.06
03/01		Withdrawal Transfer To Loan 41	288.14	591.92
03/02		Deposit ACH PAYPAL TYPE: TRANSFER ID: ZZZZ CO: PAYPAL DATA: TRANSFER	300.00	891.92
03/02		Withdrawal ACH BANNER LIFE TYPE: PREM DEBIT ID: ZZZZ CO: BANNER LIFE	19.69	872.23
03/04		Deposit ACH ELECTRONIC THEAT TYPE: DIRECT DEP ID: ZZZZ DATA: ZZZZ CO: ELECTRONIC THEAT	2054.61	2926.84
03/06		Deposit Cash	200.00	3126.84
03/09		Deposit ACH PAYPAL TYPE: TRANSFER ID: ZZZZ CO: PAYPAL DATA: TRANSFER	300.00	3426.84
03/09		Withdrawal ACH CAPITAL ONE TYPE: ONLINE PMT ID: ZZZZ CO: CAPITAL ONE	2200.00	1226.84
03/10		Withdrawal ACH AQUA AMERICA INC TYPE: Billpay CO: AQUA AMERICA INC	57.45	1169.39
03/12		Deposit ACH PAYPAL TYPE: TRANSFER ID: ZZZZ CO: PAYPAL DATA: TRANSFER	300.00	1469.39
03/16		Draft 949 Tracer ZZZZ	120.00	1349.39

statement_start_date

starting_balance

Beginning Balance/Date could stay inside a table.

Please note that **NOT** all the first item in a table is starting balance/Date.

statement_start_date and statement_end_date are mixed

In cases where the statement period (*starting_date* and *ending_date*) are combined together in a date range then the same value would be labeled twice, once for *starting_date* and one for *ending_date* as shown below:



Schwab One® Account of
NNNN

Account Number
CCCC

Statement Period
March 1-31, 2020

Transaction Detail - Purchases & Sales (continued)

Equities Activity (continued)

statement_start_date

statement_end_date



General Labeling Instructions

Occurrence Type



There are 4 occurrence types which restricts the number of annotations for each entity:

- **OPTIONAL_ONCE**: Entity can be present zero or multiple times but should have one unique value.
- **OPTIONAL_MULTIPLE**: Entity will appear zero or multiple times and with different values.
- **REQUIRED_ONCE**: Entity can be present once or multiple times but should have one unique value.
- **REQUIRED_MULTIPLE**: Entity will appear once or multiple times and with different values.

All Bank Statement form entities belongs to any of the occurrence type mentioned above .

Optical Character Recognition (OCR) Errors



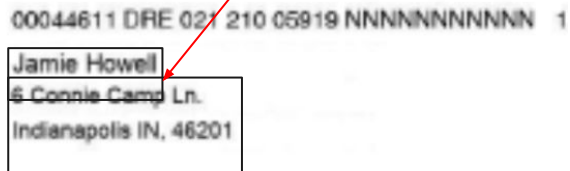
Drawing bounding boxes around entities will automatically populate a text captured through OCR (Optical Character Recognition). In some cases, the captured text is not accurate and doesn't reflect what is on the page, especially for documents that are blurry or of low resolution. These values should be corrected manually through "Value" input text.

Bounding Box Annotation

Bounding boxes should fit the correct text as much as possible:

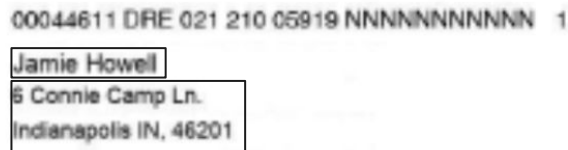
- Avoid overlapping bounding boxes. This can be tricky when values are close to each other.
- Avoid oversized boxes which incorporates background artifacts such as table borders.

Example 1



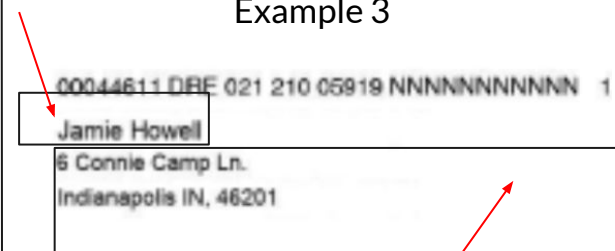
Incorrect Annotation:
Bounding boxes are overlapping

Example 2



Correct Annotation

Example 3



Incorrect Annotation:
Bounding boxes oversized

Case Sensitivity



DocAI evaluation is case sensitive and can affect the final evaluation. While annotation please be aware to fix any casing issues generated by OCR.

Please note that ENUM values are also case sensitive and the values should be fix.

Examples :

- “Yes” != “yes”
- “NO” != “No”

Money

For money entries the annotation bounding box and value text should include the currency if present on the document.

If the document contains currency with no values, then annotations could be omitted, as shown in the example on the bottom right.

Contract Summary

	This Quarter	Year To Date	Since Issue Date ¹
Beginning Value on 09/28/2018	\$201,931.64		
Total Premium	\$0.00	\$4,000.00	\$146,388.89
Total Withdrawals	\$0.00	\$0.00	\$0.00
Total Tax Withheld	\$0.00	\$0.00	
Net Change	(\$23,104.00)	(\$15,683.48)	
Ending Value ² on 12/31/2018	\$178,827.64		

FATCA filing requirement <input type="checkbox"/>	12 State tax withheld \$	13 State/Payer's state no. 000	14 State distribution \$
Date of Payment	15 Local tax withheld \$	16 Name of locality	17 Local distribution \$

www.irs.gov/form 1099r Department of the Treasury - Internal Revenue Service

