



Expense Labeling Instructions

chenblair@google.com

September 2022



General Labeling Instructions

Labeling guidance

Getting Started

- Start by annotating a small number of documents
- Review them, correct any mistakes, update the labeling guidance to make sure data is labeled consistently.

Choosing the right data/model

- A general rule is to uptrain with customer production traffic. The closer the training distribution is to production traffic, the better the performance.
- Depending on your customer's requirements, consider utilizing more than one processor in order to support multiple languages
- One model for every region (group of untrained languages) vs a universal model

Dataset Size

- Training - at least 1000 per language
- Test - at least 200 per language

Multiple Occurrence Handling

There are 2 occurrence types that we are relevant to expense labeling:

- **OPTIONAL_ONCE**: This entity is expected to have no more than **one** ground truth value.
- **OPTIONAL_MULTIPLE**: This entity can have any number of ground truth values.

For expenses, **line item** entities are **OPTIONAL_MULTIPLE**, while all other entities are **OPTIONAL_ONCE**.

Note that while **OPTIONAL_ONCE** entities can still appear multiple times with the same value. In this case, we recommend to label all appearances, except for “easy” entities (in the current entity list, only **currency**).

Optical Character Recognition (OCR) Errors

Drawing bounding boxes around entities will automatically populate a text captured through OCR. 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.

If a value doesn't exist in the document, there is no need to label it.



Expense Entity Definitions

Expense Parser Entities

In total, there are **22** entity types to be labeled.

These types can be broken down broadly into **5** categories (+ miscellaneous).

Supplier

supplier_name
supplier_address
supplier_phone

Timestamp

purchase_time
receipt_date
start_date
end_date

Line Item

line_item/amount
line_item/description
line_item/quantity
line_item/product_code
line_item/transaction_date

Amount

total_amount
total_tax_amount
net_amount
tip_amount

Payment

currency
payment_type
credit_card_last_four_digits
payment_authorization_id

Miscellaneous

traveler_name
reservation_id

Supplier Entities

A **supplier** is the business/individual who provided the goods or services involved in the expense.

supplier_name - Name of the supplier

supplier_address - Address of the supplier. In a physical transactions, this will often be the location at which the transaction was conducted.

supplier_phone - Phone number of the supplier

 Marketplace **supplier_name**
supplier_name Marketplace Sunnyvale
supplier_address 1155 Borregas Ave
Sunnyvale, CA 94089
408-555-1234 **supplier_phone**

09/09/2022 (Fri) 12:24PM
Order #123
APPR #:1000
VISA 1234

QTY	ITEM	TOTAL
2	Oven Roasted Trout	32.00
1	Herb Lentils with Anise	3.00
1	Creamy Mashed Potatoes	5.00
1	Garlic Sautéed Spinach	8.00
1	Chocolate Pot de Creme	4.00

Subtotal: \$52.00

Tax: \$4.29

Total: \$56.29

Tip: 7.80

Total: \$64.09

Timestamp Entities

Timestamp entities identify *when* a transaction occurred. Some entities only apply to *continuous* transactions that occur over the span of days such as car rental or hotel stays

purchase_time - Clock time of transaction

receipt_date - Date of transaction

start_date - Start date of a continuous transaction

end_date - End date of a continuous transaction

Car Renters

Rental Agreement	Confirmation	Car Group
123456789	0987654321	ABCD
Jack Doe		
CAR NUMBER		ABC1234 DE
RENTED:		SFO INTLAIRPT
RETURNED:		SJC INTLAIRPT
DAY OUT:	start_date	Jan 24 2022 9:14AM
DAY IN:	end_date	Jul 13 2022 1:30PM
MILES OUT:		27623
MILES IN:		29211
CAR:		Toyota Corolla
ADDITIONAL CHARGES		USD 30.24
TOTAL CHARGES		USD 158.21
PAID BY:		VISA
CREDIT CARD#:		XXXXXXXXXXXX1234
APPR#		12345
PAYMENT		USD 158.21

Thanks for renting from Car Renters! We hope to see you again soon.



Marketplace

Marketplace Sunnyvale
1155 Borregas Ave
Sunnyvale, CA 94089
408-555-1234

receipt_date 09/09/2022 (Fr) 12:24PM
Order #123

APPR #:1000
VISA 1234

QTY	ITEM	TOTAL
2	Oven Roasted Trout	32.00
1	Herb Lentils with Anise	3.00
1	Creamy Mashed Potatoes	5.00
1	Garlic Sauteed Spinach	8.00
1	Chocolate Pot de Creme	4.00

Subtotal: \$52.00

Tax: \$4.29

Total: \$56.29

Tip: 2.80

Total: \$64.09

purchase_time

Line Item Entities

Line item entities appear in the itemized list of what was purchased. There will probably be multiple of each `line_item` entity per receipt

`line_item/amount` - Total cost of line item

`line_item/description` - Description of line item

`line_item/quantity` - Quantity of line item purchased

`line_item/product_code` - Unique identifying code for line item

`line_item/transaction_date` - Date of line item purchase



Marketplace Sunnyvale
1155 Borregas Ave
Sunnyvale, CA 94089
408-555-1234

09/09/2022 (Fri) 12:24PM
Order #123

APPR #: 1000
VISA 1234

QTY	ITEM	line_item/description	TOTAL	line_item/amount
2	Oven Roasted Trout		32.00	
1	Herb Lentils with Anise		3.00	
1	Creamy Mashed Potatoes		5.00	
1	Garlic Sautéed Spinach		8.00	
1	Chocolate Pot de Crème		4.00	

Subtotal: \$52.00

Tax: \$4.29

Total: \$56.29

Tip: 7.00

Total: \$64.09

Amount Entities

Amount entities identify aggregate payment amounts owed in the transaction

total_amount - Total payment owed

total_tax_amount - Total tax owed

net_amount - Total payment, not including taxes, tips, or additional fees

tip_amount - Total tip paid



Marketplace Sunnyvale
1155 Borregas Ave
Sunnyvale, CA 94089
408-555-1234

09/09/2022 (Fri) 12:24PM
Order #123

APPR #:1000
VISA 1234

QTY	ITEM	TOTAL
2	Oven Roasted Trout	32.00
1	Herb Lentils with Anise	3.00
1	Creamy Mashed Potatoes	5.00
1	Garlic Sauteed Spinach	8.00
1	Chocolate Pot de Creme	4.00

Subtotal:	52.00	net_amount
Tax:	\$4.29	
Total:	56.29	total_tax_amount
Tip:	7.80	tip_amount
Total:	64.09	total_amount

Payment Entities

Payment entities detail the transaction payment and how it was conducted.

currency - Currency of the transaction

payment_type - What form of payment was used (VISA, AMEX, Cash, etc.)

payment_authorization_id - ID that uniquely identifies the transaction

credit_card_last_four_digits - If a credit/debit card was used, the last four digits of the card number

Car Renters		
Rental Agreement	Confirmation	Car Group
123456789	0987654321	ABCD
Jack Doe		
CAR NUMBER		ABC1234 DE
RENTED:		SFO INTL AIRPT
RETURNED:		SJC INTL AIRPT
DAY OUT:		Jan 24 2022 9:14AM
DAY IN:		Jul 13 2022 1:30PM
MILES OUT:		27623
MILES IN:		29211
CAR:		Toyota Corolla
ADDITIONAL CHARGES		USD 30.24
TOTAL CHARGES		USD 158.21 currency
PAID BY:		VISA payment_type
CREDIT CARD#:		XXXXXXXXXXXX1234 credit_card_last_four_digits
APPR#		12345 payment_authorization_id
PAYMENT		USD 158.21

Thanks for renting from Car Renters! We hope to see you again soon.

Miscellaneous Entities

traveler_name - Name of the traveler on an expense involving transportation of lodging

reservation_id - Confirmation number given to customer at time of booking

Car Renters		
Rental Agreement	Confirmation	Car Group
123456789	0987654321	ABCD
traveler_name	Jack Doe	
CAR NUMBER:		ABC1234 DE
RENTED:		SFO INTL AIRPT
RETURNED:		SJC INTL AIRPT
DAY OUT:		Jan 24 2022 9:14AM
DAY IN:		Jul 13 2022 1:30PM
MILES OUT:		27623
MILES IN:		29211
CAR:		Toyota Corolla
ADDITIONAL CHARGES		USD 30.24
TOTAL CHARGES		USD 158.21
PAID BY:		VISA
CREDIT CARD#:		XXXXXXXXXX1234
APPR#		12345
PAYMENT		USD 158.21

Thanks for renting from Car Renters! We hope to see you again soon.