Google

Invoice labeling Instructions

September 2022

Labelling 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

Google

Multi-Occurrence entities

There are 2 occurrence types that we are relevant to invoice 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 invoices, **line item and vat** 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`).

Google

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.

Google

All Entities

- invoice_id
- invoice_date
- due_date
- purchase_order
- delivery_date
- total_amount
- total_tax_amount
- net_amount
- payment_terms
- currency
- currency_exchange_rate
- carrier
- freight_amount

- line_item/description
- line_item/quantity
- line_item/product_code
- line_item/unit
- line_item/unit_price
- line_item/amount
- line_item
- line_item/purchase_order
- vat/amount
- vat/tax_rate
- vat/tax_amount
- vat/category_code
- vat

- supplier_name
- supplier_address
- supplier_email
- supplier_website
- supplier_phone
- supplier_registration
- supplier_iban
- supplier_payment_ref
- supplier_tax_id
- receiver_tax_id
- receiver_name
- receiver_address
- receiver_email
- receiver_website
- receiver_phone

- remit_to_name
- remit_to_address
- ship_to_name
- ship_to_address
- ship_from_name
- ship_from_address

Invoice Entity Definitions

Supplier entities

supplier_name

This label corresponds to the name of the organization/entity, supplying the items/issuing the invoice.Examples: "Deems Uk Ltd", "Fab India Pvt Ltd". Do not include the address/phone number in this field.

supplier_address: Full address of the organization/person/entity which has issued the invoice/supplied the items. Do not include the fax/phone number/email in this field. Example: "1345 Main Street, Chicago IL 60007"

supplier_email: Email address of the organization/person/entity which has issued the invoice/supplied the items. Example: "abc@def.com".

supplier_phone: Phone number of the organization/person/entity which has issued the invoice/supplied the items. Example: "+1 234 567 8901".

supplier_iban : This label correspond to the supplier's (entity supplying the items/issuing the invoice) IBAN (International Bank Account Number).Example: "A123 4567 8905".

Google

Supplier entities

supplier_payment_ref: Payment reference number of the organization/person/entity which has issued the invoice/supplied the items. Example: "123/4567/8905".

Alvast bedankt voor uw betaling van **71,25 euro** op rekening van:
- Liantis
- IBAN: BE37736036953428 (BIC: KREDBEBB)
- met vermelding: +++020/0090/04190+++   supplier_payment_ref

Door uw betaling binnen de 14 dagen, vermijdt u bijkomende kosten.

supplier_registration: These labels correspond to the supplier's (entity supplying the items/issuing the invoice) registration number. This is also referred as the EU VAT registration number of the supplier. Example: "BE 123 4567 8905". Associated key example: "T.V.A."

supplier_tax_id: tax id of the supplier i.e entity issuing the invoice. The tax ID format is typically composed of a 2 letter prefix followed by digits (e.g. BE0453117484)

Google

Receiver entities

receiver_name: Name of the organization/person/entity to which the invoice is made or billed. Examples: "Google Ireland Limited". Do not include the address/phone number in this field.

receiver_address: Full address of the organization/person/entity to which the invoice is made or billed. Do not include the fax/phone number/email in this field. Example: "1345 Main Street, Chicago IL 60007"

receiver_email: Email address of the organization/person/entity to which the invoice is made or billed.

receiver_website: Website of the organization/person/entity to which the invoice is made or billed.

receiver_phone: Phone number of the organization/person/entity to which the invoice is made or billed.

receiver_tax_id: The actual tax id of the customer. The tax ID format is typically composed of a 2 letter prefix followed by digits (e.g. BE0453117484)

Amount entities

total_amount: The total amount billed, including tax, in the invoice. It is the actual text indicating the monetary amount.

total_tax_amount: The correspond to the total tax amount in the invoice.

net_amount: The actual text indicating the net amount(total amount before taxes are added).

freight_amount: The actual text indicating the freight amount.

amount_paid_since_last_invoice: The actual text indicating the amount paid since last invoice.

Miscellaneous entities

ship_to_address: Full address of the organization/person/entity to which the goods are being shipped. Include City, State, Zip code & Country if specified. Do not include the fax/phone number/email in this field. Example: "1345 Main Street, Chicago IL 60007".

ship_from_address: Full address of the origin location of the shipment. Do not include the fax/phone number/email in this field. Example: "1345 Main Street, Chicago IL 60007".

remit_to_name: Full name of entity to whom the payment should be remitted to. Examples: "Google LLC".

remit_to_address: Full address where the payment should be sent to. Do not include the fax/phone number/email in this field. Examples: "1345 Main Street, Chicago IL 60007".

ship_to_name: Actual name of the entity/person, to whom the goods are being shipped to.

ship_from_name: Full name of organization/entity which is mailing the items. Do NOT confuse with the carrier/parcel delivery service.

Miscellaneous entities

carrier: This label specifies the entity/organization, mentioned in the invoice, that is being used for parcel delivery. Examples: "FedEx", "UPS", "DHL". Associated key example: "Carrier", "Shipper", "Shipment method".

currency: It is the actual currency for the amount. Examples: "USD", "GBP", "$", "€", "US $". Associated key example: "Currency", "FX".

currency_exchange_rate: It is the actual exchange rate specified in the invoice. Examples: "75 INR per USD". Associated key example: "FX rate".

invoice_id: The actual invoice ID. Include any prefix/suffix associated with the invoice id. Example- "INV-23", "123456". Associated key example: "Invoice Id", "Invoice #"

purchase_order: Purchase order is also an identifier of the transaction. Include any prefix/suffix associated with the purchase order id. Example- "PO128". Associated key example: "PO #", "Purchase Order Ref".

invoice_date: This field refers to the invoice generation date. Associated key example: "Dated:", "Invoice Date".

delivery_date: The actual delivery date mentioned in the invoice. Associated key example: "Delivery date".

due_date: The actual date when the customer should settle the invoice payment. Example: "12/31/2019", "Dec 2019", "5th May 2020". Associated key example: "Due date", "Due by", "Pay by".

payment_terms: The actual text of the payment terms. Include full text including # of days, due by date, etc (if present). Examples- "30 days", "50% due now, 50% due in 90 days". Associated key example: "Payment terms", "Terms".

Google

Line item entities

line_item/description is the description of item(s). There are usually many values in a single image. Examples: "Laptop", "Paper bundle", "HP Headphones".

line_item/product_code: Refers to items describing the product code/ID. There are usually many values in a single image, one corresponding to each item. Examples: "BN-150", "75", "B-108".

line_item/quantity: It is the actual number of units ordered. There are usually many values in a single image, one corresponding to each item. Examples: "10", "20" in "20 lbs", "30" in "30 bags".

line_item/unit: It is the measurement unit of the item being ordered. There are usually many values in a single image, one corresponding to each item. Examples: "lb", "pounds", "kg", "hours".

line_item/unit_price: It is the actual price per unit of the item being ordered. There are usually many values in a single image, one corresponding to each item. Examples: "70", "50" in "50 $/hr", "20" in "20 $/pound".

line_item/amount: The actual amount billed for a particular item in the items table. Usually, one value per line item in the items table. Examples: "20", "65.4".

line_item/purchase_order: The actual purchase order for a particular item in the items table. Usually, one value per line item in the items table. Examples: "128", "PO128", "PO 128".

Google

Line item entities

line_item: One row of a table encompassing the other line item values (line_item/description, line_item/quantity, line_item/unit, line_item/unit_price, product_id).

Example 1

T

Commercial Invoice

supplier_name **Tabcdefg Limited**

supplier_address 334 Parksideway Road, CA 32234

AIRWAY BILL NO.	INVOICE NO.	INVOICE DATE	DATE OF EXPORT
76183428	invoice_id 465720	invoice_date 04/28/2020	04/08/2022

EXPORTER/ SHIPPER

COMPANY NAME — Asdfgh Logistics ship_from_name

ADDRESS — 162 Mixen Avenue Alberta, AB 88225 ship_from_address

CONTACT NAME — John Doe

PHONE/ FAX — 987-987-9876

EMAIL — manager@asdfg.com

COUNTRY OF EXPORT — Canada

SHIP TO / COSIGNEE

COMPANY NAME — Abcxyz Traders ship_to_name

ADDRESS — 45 Lightning Road Arizona, AZ 88776 ship_to_address

CONTACT NAME

PHONE/ FAX — 321-321-1234

EMAIL — proprietor@abcxyz.com

Product	Qty	Unit Price	Amount
Optical Mouse	45	$230.00	$10,350.00
line_item/description	line_item/quantity	line_item/unit_price	line_item/amount
Monitor 14 Inch	23	$100.00	$2,300.00
line_item/description	line_item/quantity	line_item/unit_price	line_item/amount
Dot Matrix Printer	5	$125.00	$625.00
	line_item/quantity	line_item/unit_price	line_item/amount

Sub Total net_amount $13,275.00

Discount $210.00

Tax total_tax_amount $80.00

Total Value currency $13,145.00

total_amount — total_amount

line_item (×3)

Example 2

Commercial Invoice

supplier_name **Tabcdefg Limited**

supplier_address 334 Parksideway Road, CA 32234

AIRWAY BILL NO.	INVOICE NO.	INVOICE DATE	DATE OF EXPORT
76183428	invoice_id 465720	invoice_date 04/28/2020	04/08/2022

EXPORTER/ SHIPPER

COMPANY NAME   Asdfgh Logistics ship_from_name

ADDRESS   162 Mixen Avenue Alberta, AB 88225 ship_from_address

CONTACT NAME   John Doe

PHONE/ FAX   987-987-9876

EMAIL   manager@asdfg.com

COUNTRY OF EXPORT   Canada

SHIP TO / COSIGNEE

COMPANY NAME   Abcxyz Traders ship_to_name

ADDRESS   45 Lightning Road Arizona, AZ 88776 ship_to_address

CONTACT NAME

PHONE/ FAX   321-321-1234

EMAIL   proprietor@abcxyz.com

Product	Qty	Unit Price	Amount
line_item **Optical Mouse** line_item/description	45 line_item/quantity	$230.00 line_item/unit_price	$10,350.00 line_item/amount
line_item **Monitor 14 Inch** line_item/description	23 line_item/quantity	$100.00 line_item/unit_price	$2,300.00 line_item/amount
line_item **Dot Matrix Printer**	5 line_item/quantity	$125.00 line_item/unit_price	$625.00 line_item/amount

Sub Total   net_amount $13,275.00

Discount $210.00

Tax   total_tax_amount $80.00

Total Value   currency $ total_amount 13,145.00   total_amount

Google