



Purchase Order labeling Instructions

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 up-train 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

Multi-Occurrence entities

There are 2 occurrence types that we are relevant to purchase order 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 Purchase Order, **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.



Purchase Order Entity Definitions

All Entities

- currency
- delivery_date
- payment_terms
- purchase_order_date
- purchase_order_id
- receiver_name
- ship_to_address
- ship_to_name
- total_amount

- line_item/amount
- line_item/description
- line_item/product_code
- line_item/purchase_order_id
- line_item/quantity
- line_item/receiver_reference
- line_item/unit_of_measure
- line_item/unit_price

Receiver entities

receiver_name: Name of the organization/person/entity to which the purchase order is made or billed. Examples: "Zinfra".

ship_to_address: These labels correspond to the entity/organization address to whom the items are being shipped to. The key for this field might be the same as the ship_to_address field. Example : ship to , deliver to.

ship_to_name: Actual name of the entity/person, to whom the goods are being shipped to. Include full organization name including prefixes such as "LLC", "Pvt Ltd". Example: Google LLC

Line item entities

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

line_item/product_code: It 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/purchase_order_id: Actual purchase order identifier/number. Include any prefix/suffix associated with the purchase order id. Example- "PO128".

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/receiver_reference: It refers to customer code assigned by the receiver to the product or service that is being bought. For example, a supplier may have a product code XYZ, but the receiver has defined the same product in their system with code ABC.

line_item/unit_of_measure: 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 \$/hr", "20 \$/pound".

Amount entities

currency: It refers to text specifically declaring the currency used in the Purchase order. Examples: "USD", "GBP", "\$", "€", "US \$"

total_amount: The actual text indicating the monetary amount (total amount after taxes are added). Examples: "343.44", "8,372.81".

Miscellaneous entities

delivery_date: This field refers to the delivery date for the products and services being ordered. Example: “Delivery date”

payment_terms: The actual text of the payment terms. Include full text including # of days, due by date, etc (if present). Examples- “NET 30 days”, “NET 45”.

purchase_order_date: This field refers to the purchase order date. In some cases, only month & year may be mentioned, instead of the full date. Example “Dec 2019”. Please consider such values as valid and annotate them. Don’t include words like “Date” or “P/O Date”.

purchase_order_id: Purchase order is also an identifier of the transaction. Include any prefix/suffix associated with the purchase order id. Example- “P0128”.

Example 1:

receiver_name



Abcxyz Traders

PO Acknowledgment

PO Type: Manual Page 1 of 2
 Date/Time: 2021-10-21 13:10:49
 Sales Order: 1281982888

purchase_order_id

Sales Order: 1281982888 PO#: 234532
 Freight Order: Cust Ref Num: 2838723
 Order Date: 12/02/21 Ship Location: 7719 - CINCINNATI INTERSTATE
 Pricing Date: 12/02/21 Shipping Method: UPS COLLECT
 Req Delv Date: 12/5/21 Order Type: Standard
 Ship Date: 12/5/21 Payment Terms: Net 30
 Currency: USD Primary Broker: 1299188882 - NO BROKER

Sold-To 12717272772 **Ship-To** 87287237378
 ZXC INDUSTRIES
 45 LIGHTNING ROAD
 ARIZONA AZ 88776-8328
 Country: US Country: US

Purchase_order_date

delivery_date

currency

payment_terms

ship_to_name

ship_to_address

line_item/product_code * 10

line_item/quantity * 9

line_item/description * 10

line_item/receiver_reference * 10

line_item/amount * 9

line_item/unit_price * 9

line_item/unit_of_measure * 9

Item	Material	Customer Material Number	GTIN	Product Description	Order Qty	Est Order Net Weight	Pricing Units	Price/Unit	Cust Exp Price	Price/Allowances/Charges	Est. Extended Amount
10	20200003323	21212124	00237273236746	40/4 FLT BEEF ONK/FRM MAR ASSOL	160 CS	1,600.00 LB	160	\$46.43000 CS	List Price DEL-CB	\$7,428.80	\$7,428.80
20	20200001235	0003728327373	00233737377373	72/4 FC BEEF HOAGIE SOY	12 CS	216.00 LB	12	\$115.15000 CS	Published Prc DL-CB	\$1,381.80	\$1,381.80
30	20200001244	0002372377334	00372837239237	CK CHICKEN BREAST 3.00	45 CS	506.25 LB	45	\$48.71000 CS	Published Prc DL-CB	\$2,191.95	\$2,191.95
40	20200001266	0000238723773	00072374737348	4/2.5# FC MAR RDSK MEAT	100 CS	1,000.00 LB	100	\$69.10000 CS	Price OI Net - DEL	\$6,910.00	\$6,910.00
50	20200008488	0000372323674	00023238382823	40/4 PAST BA BF SIR NMAR	30 CS	300.00 LB	30	\$69.21000 CS	Published Prc DL-CB	\$2,076.30	\$2,076.30
60	20200008389	0066236626363	00712781218278	48/4 BA BF SIR MAR	100 CS	1,200.00 LB	100	\$80.96000 CS	Published Prc DL-CB	\$8,096.00	\$8,096.00
70	20200008388	0002382377358	00128128771277	27/6 BA BEEF MARINATED	20 CS	200.00 LB	20	\$47.90000 CS	Price OI Net - DEL	\$958.00	\$958.00
80	20200002982	0006732677324	00128812877124	40/4 FLT BEEF SIRLOIN MAR SEZEI	20 CS	200.00 LB	20	\$70.75000 CS	Published Prc DL-CB	\$1,415.00	\$1,415.00
90	20200004839	0003534535446	00128912128828	40/4 FLT CHK BRST MAR ROMA	10 CS	100.00 LB	10	\$48.59000 CS	List Price DEL-CB	\$485.90	\$485.90
100	20200002893	0182891828912	20200002893	CONSTRAINED INVENTORY							

Example 1:



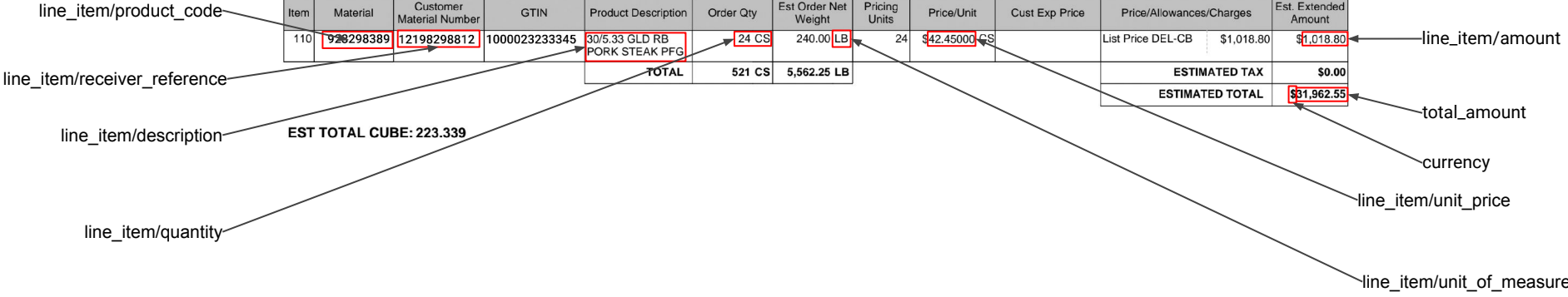
Abcxyz Traders

PO Acknowledgment

PO Type: Manual Page 2 of 2
 Date/Time: 2021-10-21 13:10:49
 Sales Order: 1291982888

Item	Material	Customer Material Number	GTIN	Product Description	Order Qty	Est Order Net Weight	Pricing Units	Price/Unit	Cust Exp Price	Price/Allowances/Charges	Est. Extended Amount
110	928298389	12198298812	1000023233345	30/5.33 GLD RB PORK STEAK PFG	24 CS	240.00 LB	24	\$42.45000		List Price DEL-CB \$1,018.80	\$1,018.80
TOTAL					521 CS	5,562.25 LB				ESTIMATED TAX	\$0.00
ESTIMATED TOTAL											\$31,962.55

EST TOTAL CUBE: 223.339



Feedback

1. Should Sales Order be invoice_id?
2. Receiver_name should be the “ZXCV INDUSTRIES” under “Sold-To”, not the title.
3. GTIN should be line_item/product_code, not Material.
4. All the comments on page 1 also applies to page 2.