

These are instructions for entity labeling

You will be asked to label subset of entities:

<ul style="list-style-type: none">○ invoice_id○ invoice_date○ due_date○ purchase_order○ delivery_date○ total_amount○ total_tax_amount○ net_amount○ payment_terms○ currency○ currency_exchange_rate○ carrier○ amount_paid_since_last_invoice○ freight_amount	<ul style="list-style-type: none">○ 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/total_amount (new field)○ vat/category_code○ vat	<ul style="list-style-type: none">○ 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	<ul style="list-style-type: none">○ remit_to_name○ remit_to_address○ ship_to_name○ ship_to_address○ ship_from_name○ ship_from_address
--	---	--	--

Special notice

- You don't need to annotate any `entity_type:key`, and for `entity_type:key` you don't need to delete existing ones or fix them either.
- But for `entity_type value`, you **must** delete invalid existing ones and fix them.
- Except for `line_item` and `vat`, other entities should not have any overlap with another in terms of text span. The bounding box may overlap.
- You need to fix the text if it's not correct.
- `entity_type:key` has been removed in this instruction, you only need to annotate values. There is no need to remove existing `entity_type:key` though you can choose to do if it helps you with annotating values.

Task Type 0

Classify the invoice type

Goal: classify the type of invoice document

You will be presented with a drop down where you can select multiple `invoice_type` options. Since these are invoices, you should choose only 1 of 3 options (i.e. ignore other drop down options):

`invoice_statement`

This is a regular invoice and the most common type of invoice.

`credit_note`

This invoice type provides an adjustment for a prior invoice (`invoice_statement`) by providing a credit to the customer's account, which reduces the amount the recipient needs to pay to the supplier.

`debit_note`

This invoice type also provides an adjustment for a prior invoice, but it instead debits the customer for an additional charge

The screenshot shows the DataCompute interface for a classification task. At the top, the DataCompute logo is visible. Below it, a progress indicator shows a timer at 37:51 and a step counter at 1. The task name is "invoice-sandbox-us-processor-P1". There are two tabs: "ENTITY" and "CLASSIFY", with "CLASSIFY" being the active tab. The instruction reads: "Select the options below to classify this document." Below this, there is a dropdown menu for the "invoice_type" field. The dropdown is open, showing three options: "invoice_statement" (which is highlighted), "credit_note", and "debit_note".

Task Type 1

Label entities

Goal: draw bounding boxes for classes in the label set

You may be asked to label value-only entities.

Some of the entities have an associated key (e.g. "Remit To" / "Invoice ID"), some might not. You should annotate as long as they are valid, with/without keys.

IMPORTANT: Please also correct the text if the OCR results are not correct. You must correct punctuations, capitalization, e.g:

- Actual text is "ABC" but the extracted text is "abc"
- Actual text is "123" but the extracted text is "\$123"
- Actual text is "12.34" but the extracted text is "12,34"

See next slides for more examples.

INVOICE 

Date: July 27th 2017
Invoice # 0480

Deerns UK Ltd
Pentagon House
52-54 Southwark Street
London
SE1 1UN
Phone: +44 (0)20 3690 6990
Jinal.patel@deerns.com

MAIL TO Attn: Accounts Payable
Iron Mountain Slovakia s.r.o
Account N/O 221B
Na Pántoch 18
831 06 Bratislava, LC Rača
Slovakia

BILL TO **Google Ireland Limited**
50723, 1st and 2nd Floor
Gordon House
Barrow Street
Dublin 4
Ireland

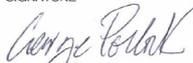
CLIENT PURCHASE ORDER NUMBER	DEERNS PROJECT NO	PAYMENT TERMS
527647	RGB400.000052.00 - 0002	30 days

DESCRIPTION	LINE TOTAL (€)
GRQ WABO File Structure WAF - July 17	9,982.00

PROJECT SUMMARY

Total Order	€ 100,000.00
Previous Invoice	€ 88,758.40
Present Invoice	€ 9,982.00
Remainder	€ 1,259.60

SUBTOTAL	9,982.00
VAT (0%)	
REVERSE CHARGE	0
TOTAL	9,982.00

SIGNATURE 

GEORGE POLLOCK
MANAGING DIRECTOR
DEERNS UK LTD

Please submit payments to:

Account Name	Deerns UK Ltd
Account Number	0179574140
Account IBAN	NL85RAB00179574140
Account SWIFT Code	RABONL2U

Deerns UK Ltd
Pentagon House
52-54 Southwark Street
London, SE1 1UN

Company Number 7459589 VAT Registration Number GB 110 7425 57

Mention text examples

Label entities

Here are some examples where mention text must be fixed manually.

The problematic part is circled in red.

8.5 x 11 Colored
C2S00 Print

Type: line_item/description
Gt Text: 8.5 x 11 Colored C2\$00 Print

Wrong letter case

AMOUNT DUE \$9,900

Type: amount_due
Gt Text: \$9,900

Extra character

PRICE

\$55.22

\$29.63

\$57.64

Type: line_item/amount
Gt Text: 37.64

Wrong punctuation

Repeat the step for different labeling tasks

Example 1

supplier_name → Acme Pest Control LLC
supplier_address → 145 Corporate Ln
supplier_phone → 555-314-1888
receiver_name → Carrie Webb
receiver_address → 909 Kuvalls Fort Apt. 525
payment_terms → Net 30
purchase_order → CC-2342
line_item → 3 Pest control chemical
line_item → 6 First Strike Soft Bait
line_item → 4 Labor
due_date → 12/13/2019
net_amount → \$331.00
total_tax_amount → \$29.79
total_amount → \$372.24
lang_en → []

Invoice
Acme Pest Control #126935
11/27/2019

Bill To: Carrie Webb
909 Kuvalls Fort Apt. 525
Bannockburn, IL 60015
United States

Service Location: Carrie Webb
909 Kuvalls Fort Apt. 525
Bannockburn, IL 60015
United States

Terms	Due Date	Sales Team
Net 30	12/13/2019	Amy 13 - Nicole 13-3

PO #	Created From	Job
CC-2342	Sales Order #2026201	Job #33508

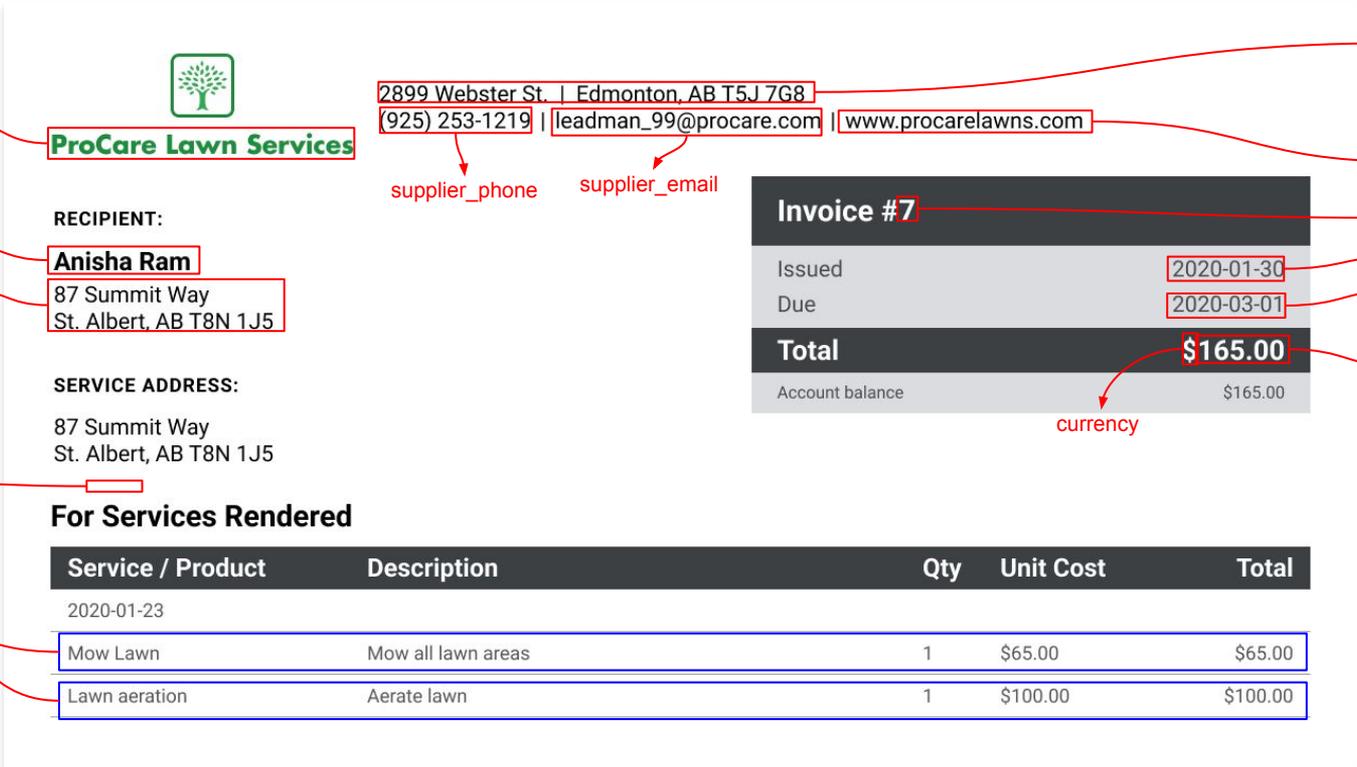
Quantity	Item	Unit Price	Amount
3	Pest control chemical	\$45.00	\$135.00
6	First Strike Soft Bait	\$3.50	\$28.00
4	Labor	\$42.00	\$168.00

Subtotal	\$331.00
Tax (%)	\$29.79
Service fee	\$11.45
Total	\$372.24
Amount Paid	\$0.00
Amount Due	\$372.24

126935

1 of 1

Example 2



The image shows an invoice from ProCare Lawn Services to Anisha Ram. The invoice includes the supplier's contact information, the recipient's name and address, the service address, and a table of services rendered. A summary table shows the total amount of \$165.00. Red arrows point from labels to specific fields in the invoice.

Supplier Information:

- supplier_name: ProCare Lawn Services
- supplier_address: 2899 Webster St. | Edmonton, AB T5J 7G8
- supplier_phone: (925) 253-1219
- supplier_email: leadman_99@procare.com
- supplier_website: www.procarelawns.com

Recipient Information:

- receiver_name: Anisha Ram
- receiver_address: 87 Summit Way, St. Albert, AB T8N 1J5

Service Address:

87 Summit Way
St. Albert, AB T8N 1J5

lang_en: []

Invoice Summary:

Invoice #	7
Issued	2020-01-30
Due	2020-03-01
Total	\$165.00
Account balance	\$165.00

currency: \$

For Services Rendered

Service / Product	Description	Qty	Unit Cost	Total
2020-01-23				
Mow Lawn	Mow all lawn areas	1	\$65.00	\$65.00
Lawn aeration	Aerate lawn	1	\$100.00	\$100.00

line_item: []

Repeat for different examples

supplier_name

These labels correspond to the name of the organization/entity, supplying the items/issuing the invoice.

supplier_name: Name of the organization/person/entity which has generated the invoice/supplied the items.

Examples: “Deems Uk Ltd”, “Fab India Pvt Ltd”.

Do not include the address/phone number in this field.

Specific Instructions:

1. In some cases, supplier information is included in the header without identifying keys (Example 1)
2. Include full organization name including suffixes such as “[LLC](#)”, “[Pvt Ltd](#)”
3. You may annotate more than one bounding box that belong to **supplier_name**. For example, [screenshot](#), annotate both the text under logo and right to logo as **supplier_name**
4. For logos
 - a. If [mention_text](#) returns human understandable supplier name from the LOGO, label it
 - b. If it is mixed unreadable text and subtitle, only label the name
 - c. If it is hard to decide whether the LOGO supplier name should be annotated and there is supplier name clearly printed, skip the LOGO

Logo examples

Supplier name

Here are examples of when to and when not to label a logo.



Do

The logo's text is legible and readable by mention_text.



Don't

The mention_text would be broken by "SIS".

Repeat for other fields