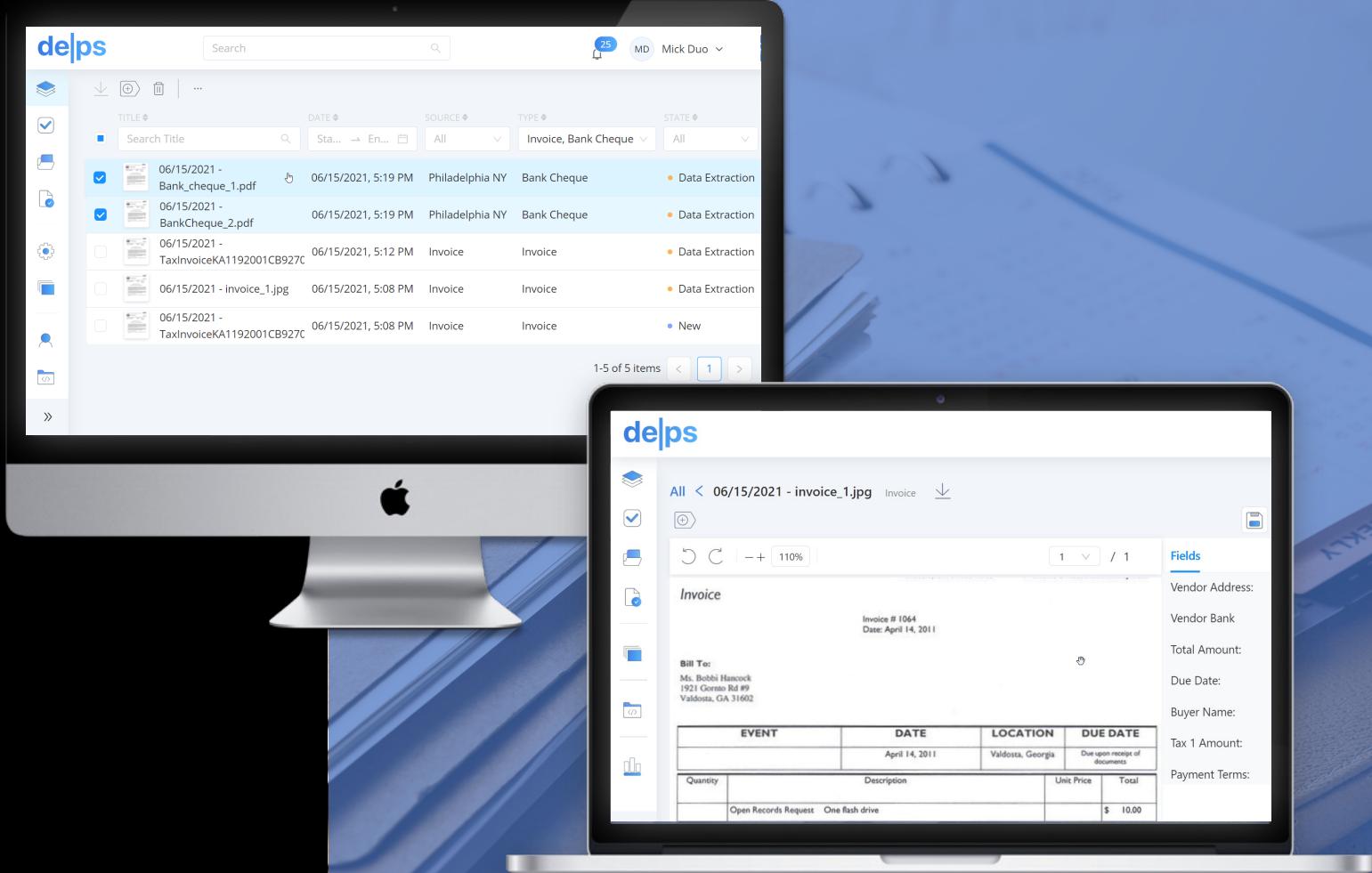


Labeling Tool

User Guide

2022
Version 2



Content

-
- 1 Labeling Tool Overview**
 - 2 Labeling Process Workflow**
 - 3 Tables Labeling**
 - 4 Key-value Pair Labeling**

Labeling Tool Overview

Data labeling is the act of adding keywords to unstructured data, like text, images, so that machines can learn to automatically recognize the concepts that these keywords describe.

Data labeling is the first step in the Artificial Intelligence lifecycle. It is the basis for developing accurate predictive models that yield a wide variety of results. The labeled data is used to train AI models and create document templates.

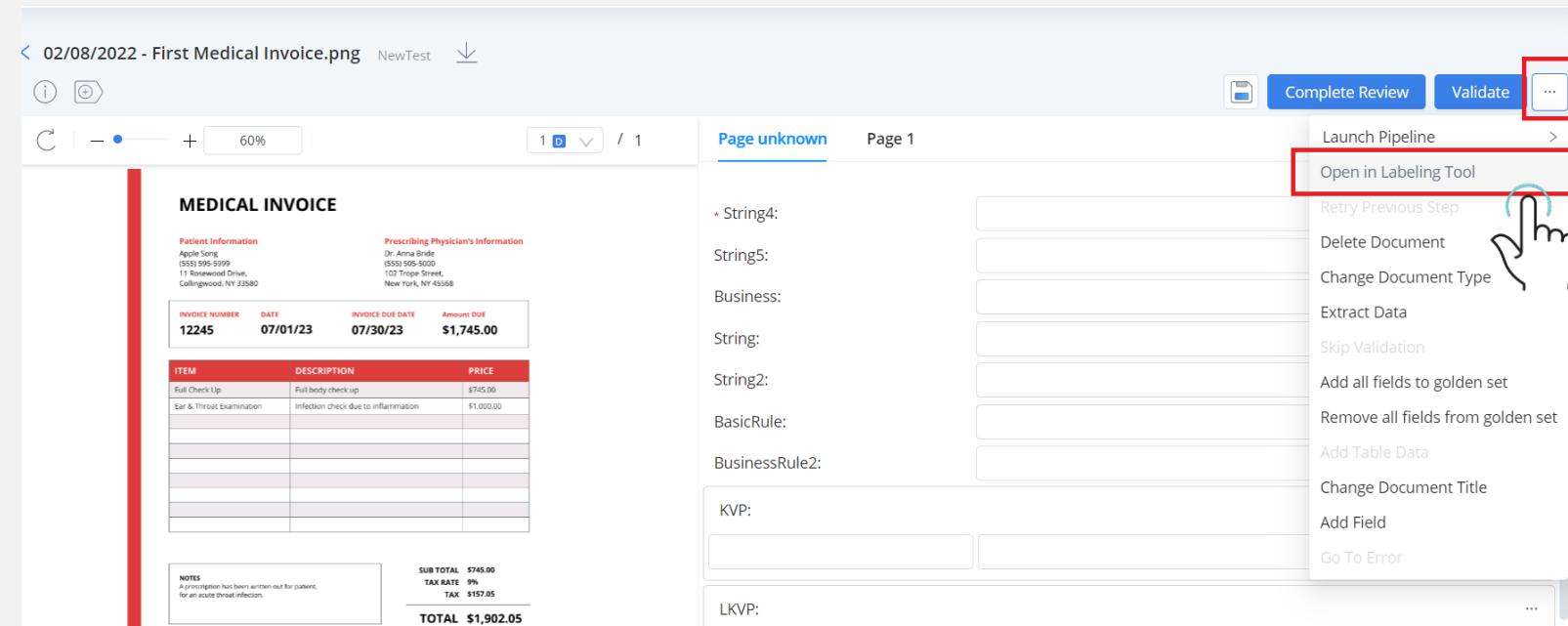
Labeling tool is used for editing table layout and labeling in the document. The user applies a labeling tool to label the data in the documents for proper extraction.

The applying of the data labeling tool will improve the process of extracting data from documents and preserve the data structure of the text, tables, images, as well as highlight only the necessary parts when extracting data.

TIP

User can open Labeling tool via “Open in Labeling Tool” action available in the upper-right corner of right pane (the document opens in a separate window for labeling).

This action is available only in "In Review" state.



Labeling page view

Icon	Description
	Main menu: allows user to import / export markup and shows information about hotkeys
	Save: saves user changes to the markup and sends marked data to extraction
	Undo: allows user to undo previous action
	Pointer: selects and moves selected markup
	Drag: moves canvas
	Label: allows user to mark up text on a page and switch to the "Label"/ "Table" labeling tool
	Detect table feature
	Merge/split table cells
	Delete: allows user to delete the markup This button is available only if a Pointer is selected.

Main menu and features

Document scaling

Properties / Assign to tab

The screenshot displays the 'Labeling page view' interface with several floating toolbars and panels:

- Main menu and features:** Shows icons for Import markup, Export markup, and Hot keys info.
- Document scaling:** Includes zoom controls (-, +, 100%) and a status bar showing page numbers (1 / 1).
- Properties / Assign to tab:** Shows a preview of a 'Proforma Invoice' document with extracted fields: Acme Industries (SHIP TO), Globex Corporation (BENEFICIARY), PROFORMA INVOICE # 000231, DATE 11/05/2020, and DUE DATE 11/05/2020.
- Hot Keys:** A detailed panel listing various hot keys with their keyboard shortcuts:

Action	Hot Key
Split whole row/column	Shift + Left mouse button
Table tool	Shift + T
Delete	Del
Undo all actions	Ctrl + Alt + Z
Collapse/expand side bar	Ctrl + Backslash
Detect tables	Shift + D
Merge tool	Shift + M
Undo last action	Ctrl + Z
Copy	Ctrl + C
Drag	Alt + Left mouse button
Label tool	Shift + L
Split tool	Shift + K
Redo last action	Ctrl + Y
Paste	Ctrl + V
Recognize	Shift + R
- All Objects / JSON Preview:** A panel showing the JSON structure of the invoice data.

Labeling process workflow

How to start labeling documents

1. The user uploads the document by specifying the document type and checking the box «Send to extraction».

Upload Documents

Invoice

Select Processing Type

Select Engine

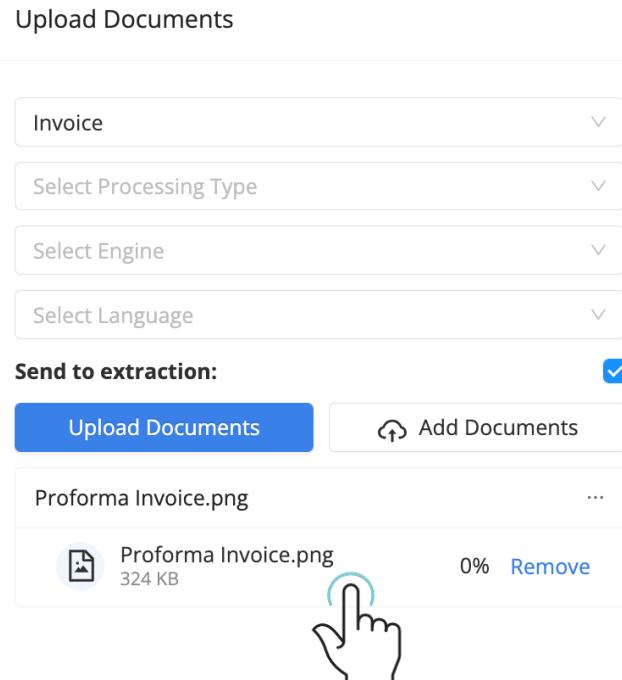
Select Language

Send to extraction:

Upload Documents Add Documents

Proforma Invoice.png

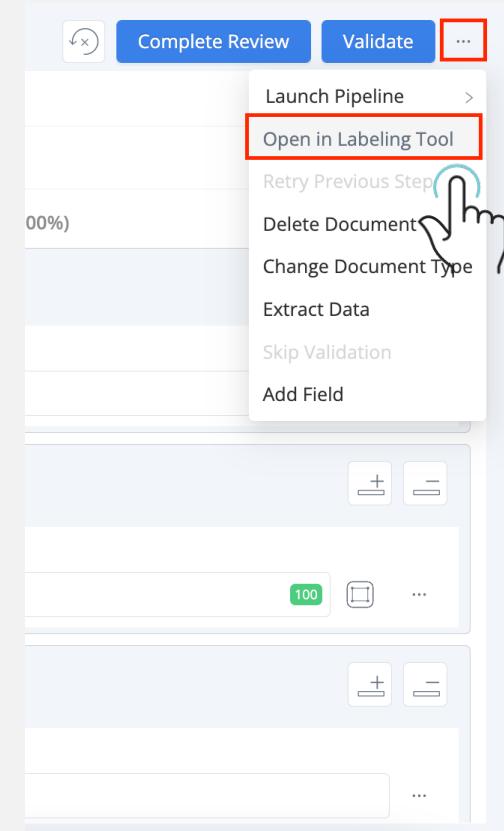
Proforma Invoice.png 324 KB 0% Remove



2. User opens the document on Storage tab and clicks the «Start Review» button.



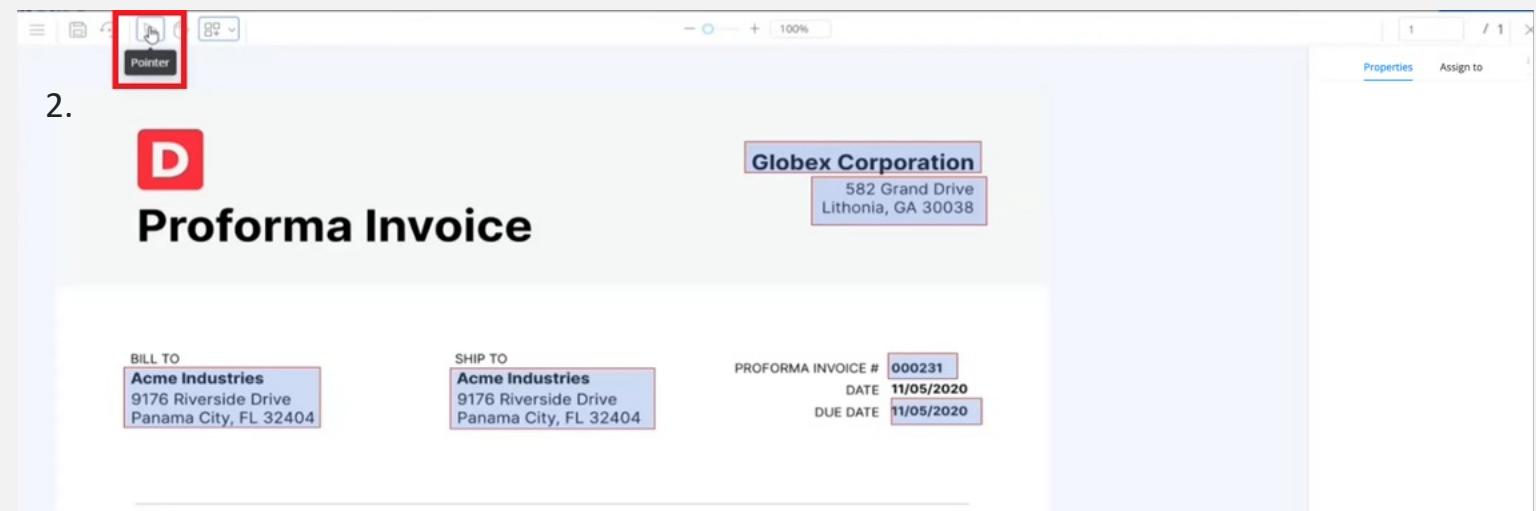
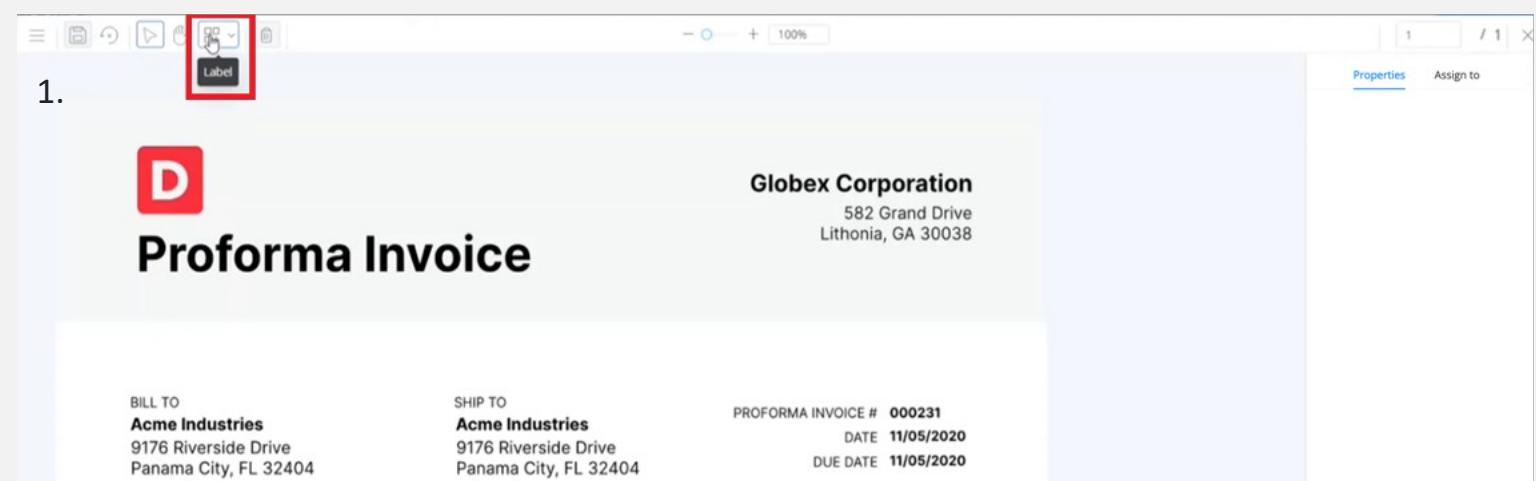
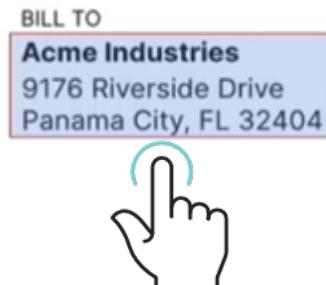
3. The Review starts, and the User can select the «Open in Labeling Tool» action from the action bar «...».



How to make labels

Step 1. Open the document in the Labeling Tool and choose the «Label» action to draw labels in the document. The user can mark a single word/number, or several words, numbers and lines in one label.

Step 2. To draw a label, the user applies the cursor and the left mouse button. By holding down the left mouse button, the user selects an area to create a new label. When the label area is highlighted, the system shows it in blue with a red outline.

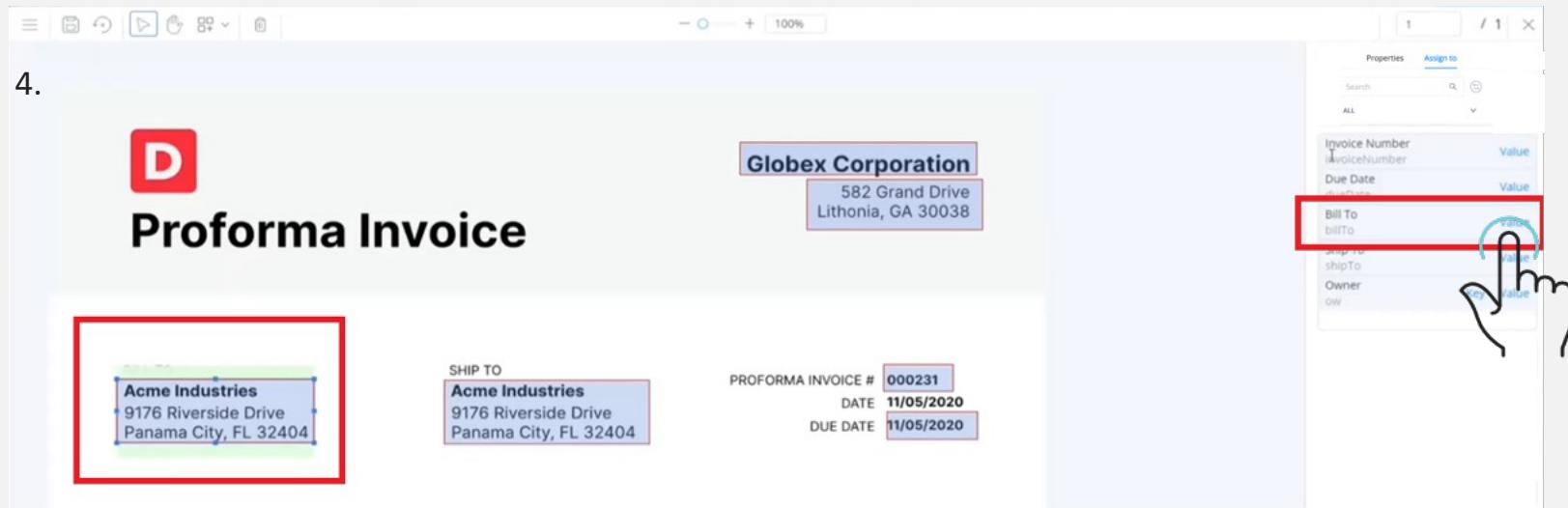
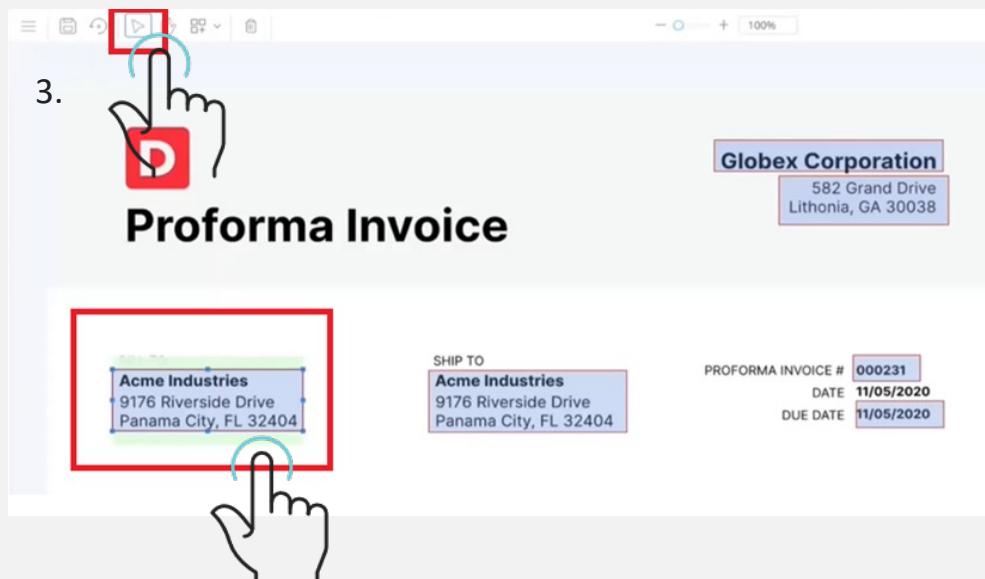


How to assign labels

Step 3. After a new label is drawn, the user selects the «Pointer» action and left-clicks on the label area. Then the label is highlighted in green at the top and bottom. This means that the label is ready for assignment.



Step 4. Now when the label is selected by the «Pointer», the list of fields available in this document type is displayed in the “Assign to” tab on the right side. The user assigns the selected label to one of the fields on the right by left-clicking on the word "Value" opposite the field name.



Labels are assigned

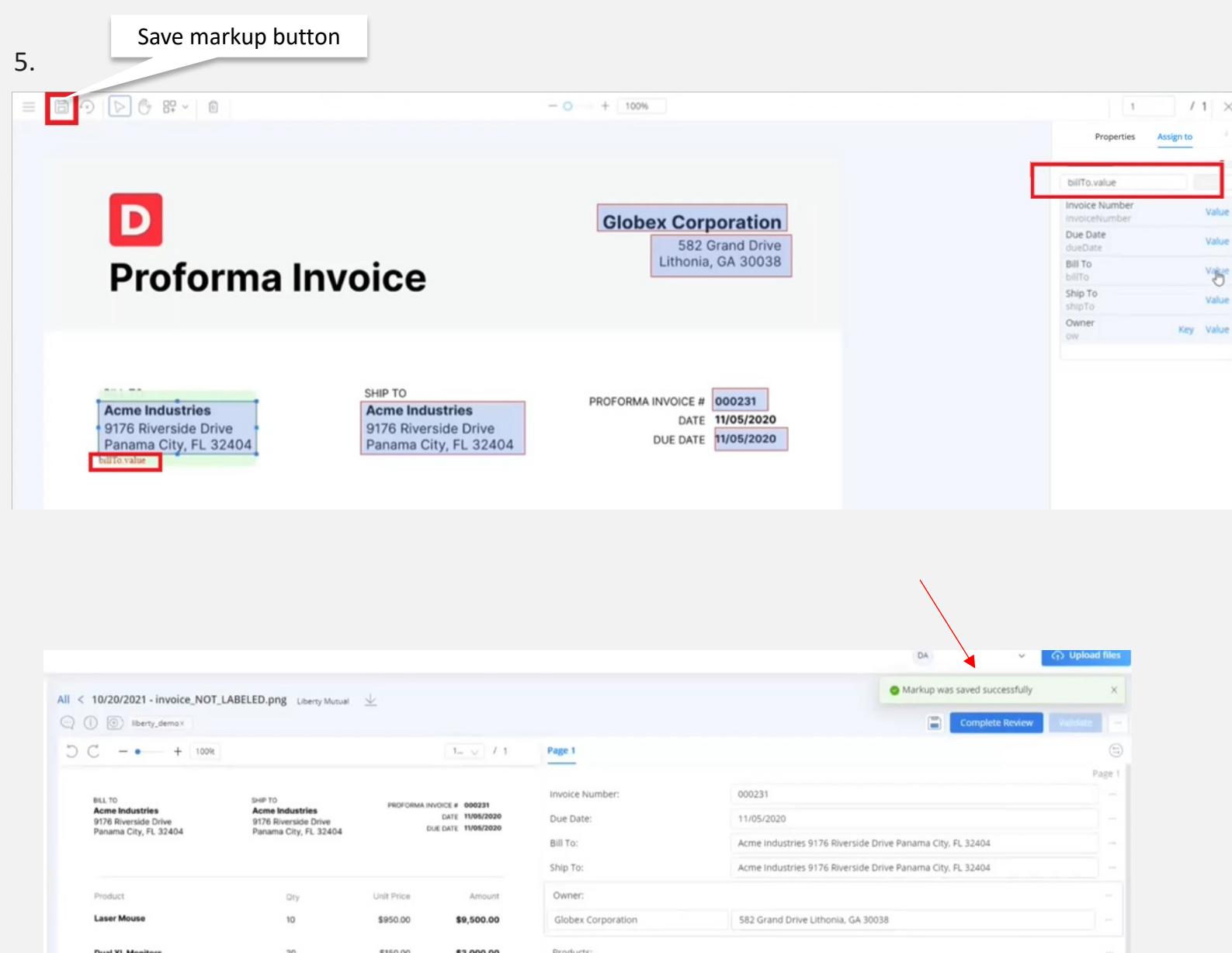
Step 5. The label now belongs to a specific field. The system assigns the field name to the label and displays it at the bottom of the label (red font). Similarly, the user assigns labels to other fields.

In the processes of drawing labels, the user can save the markup by clicking the "Save" button on the top bar. The system displays the notification "Markup was saved successfully".

The user has three saving options:

- save without extraction – saves markup, user can continue work in Labeling Tool
- save & extract – saves markup, extracts data, user can continue work in Labeling Tool
- save, extract & close – saves markup, extracts data and closes Labeling Tool, user is redirected to Document Review page

If necessary, the user can re-label the data in the document by repeating the labeling process and saving the results. The system allows making changes to the markup.



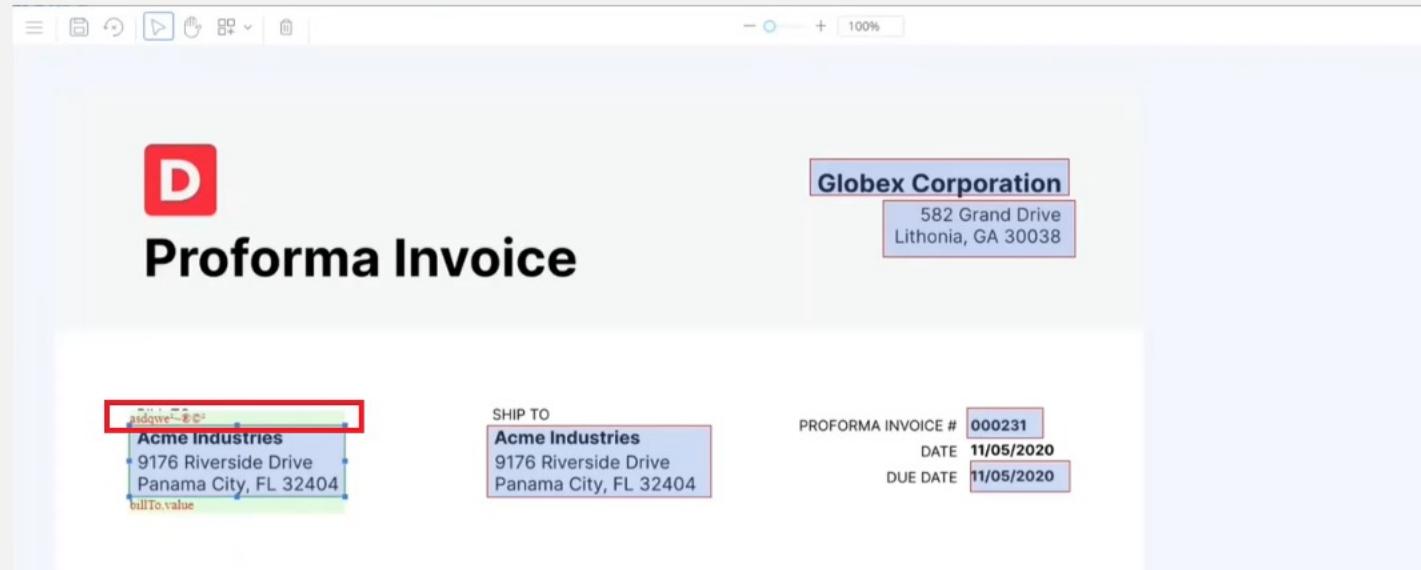
Labeling features

Extracted content

The user can create, modify and delete extracted content for each label.

The extracted content will appear at the top of the highlighted label (red font). The system also supports adding special characters to extracted content.

Show special characters



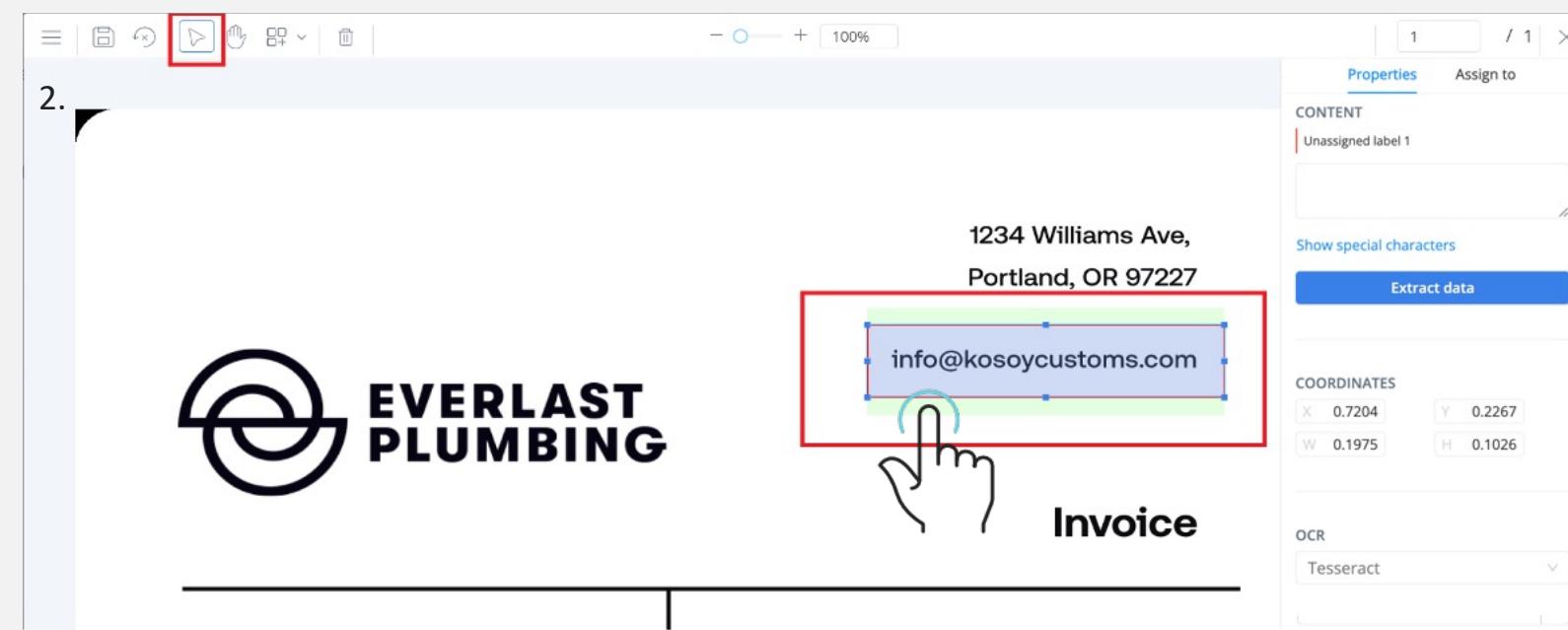
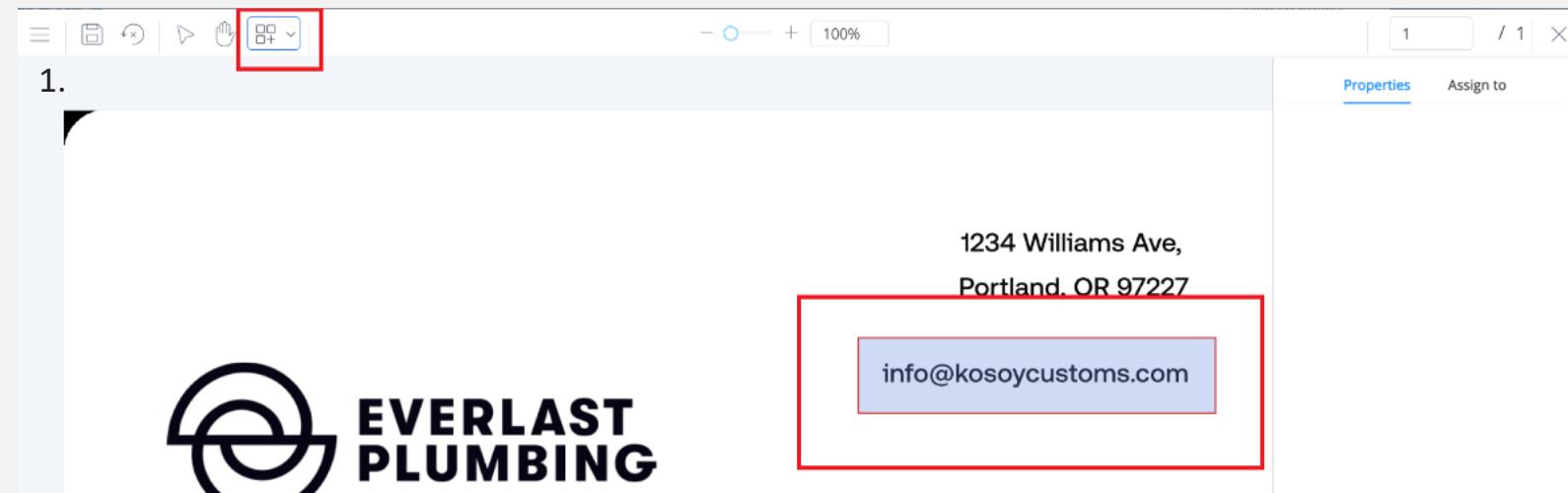
Labeling features

Direct extraction

The user can send data to be extracted directly from the Labeling Tool.

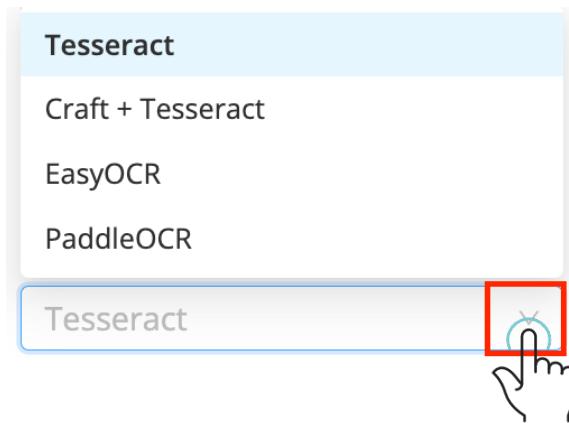
Step 1. Select the «Label» action to draw labels in the document. By holding down the left mouse button, the user selects an area to create a label. When the label area is highlighted, the system shows it in blue with a red outline.

Step 2. Select the "Pointer" action and left-click on the label area, the label is highlighted in green.

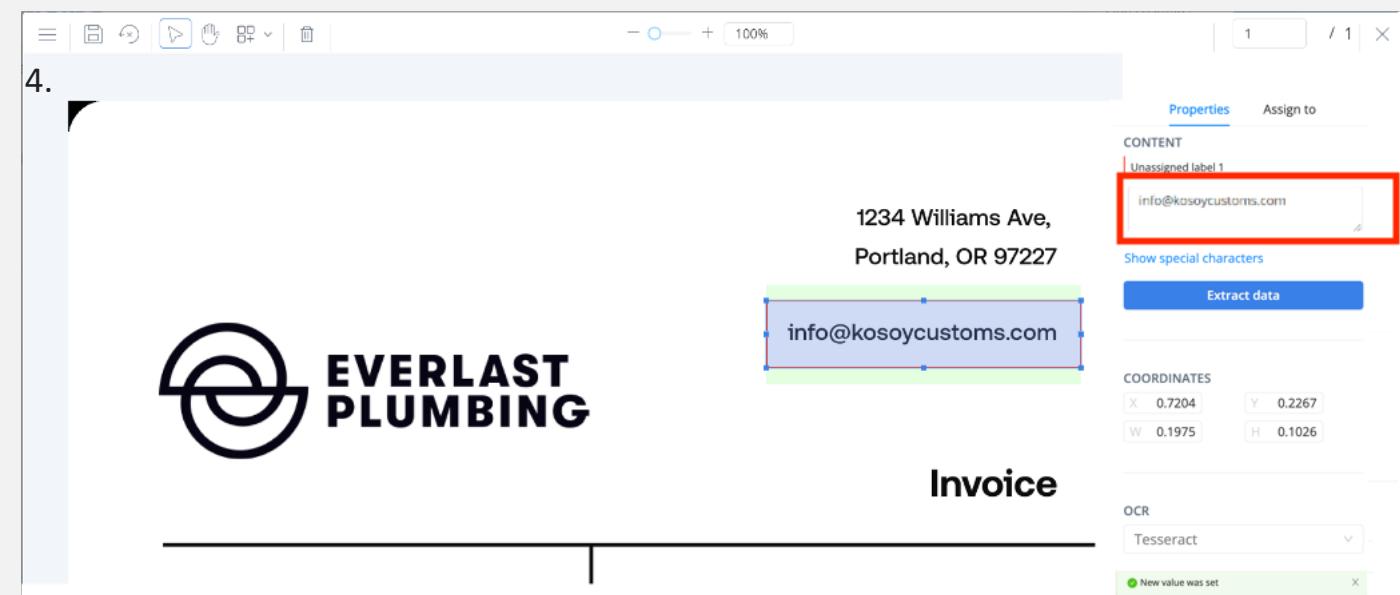
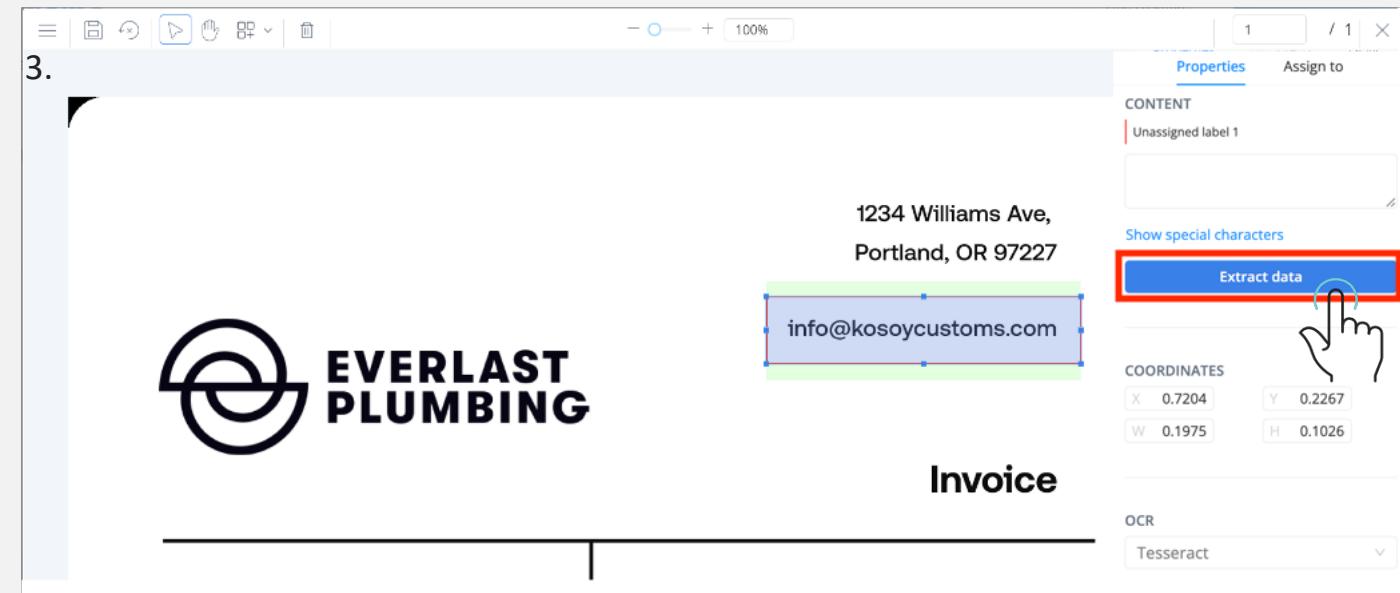


Labeling features

Step 3. Click the «Extract data» button to extract data from the selected label. The user can select the engine for extraction by clicking on OCR dropdown. By default, the Tesseract engine is applied.



Step 4. The data is extracted and displayed in the box. In this case, the system displays the notification «New value was set».



Tables Labeling

Step 1. The user uploads the document by specifying the document type and checking the box «Send to extraction».

Step 2. The user opens the document in Storage tab and clicks the «Start Review» button.

Step 3. The Review starts, and the User can select the «Open with Labeling Tool» action from the action bar «...».

Step 4. The user should switch from the «Label» menu to «Table» and activate table labeling features:

 - detect tables

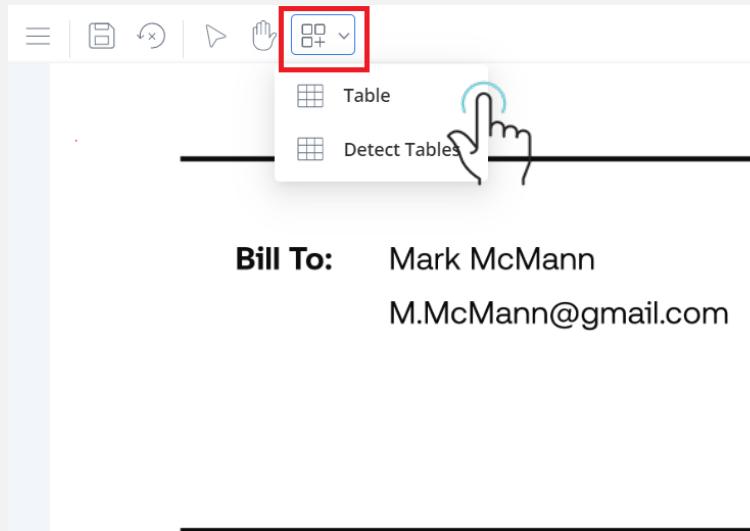
 - merge/split cells

TIP

There are two ways to markup tables in the Labeling Tool:

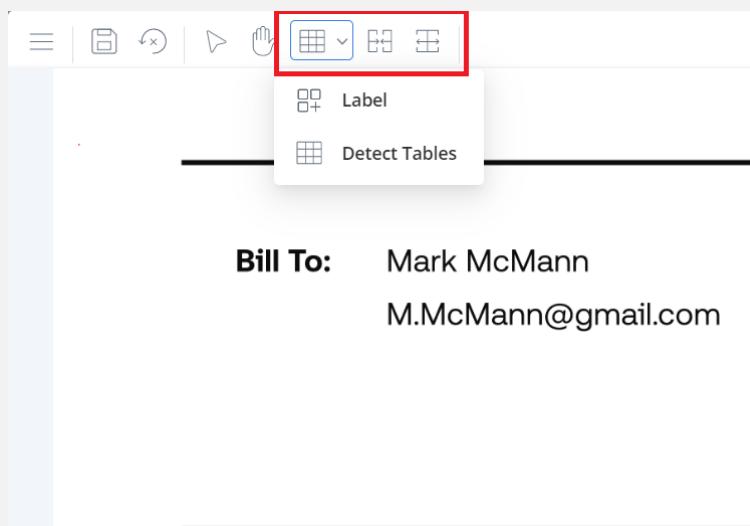
- automatic detect tables
- manual table markup

4.1



>> The user clicks the «Label» button, an action menu appears, in which he should select the «Table» or «Detect Tables» action.

4.2

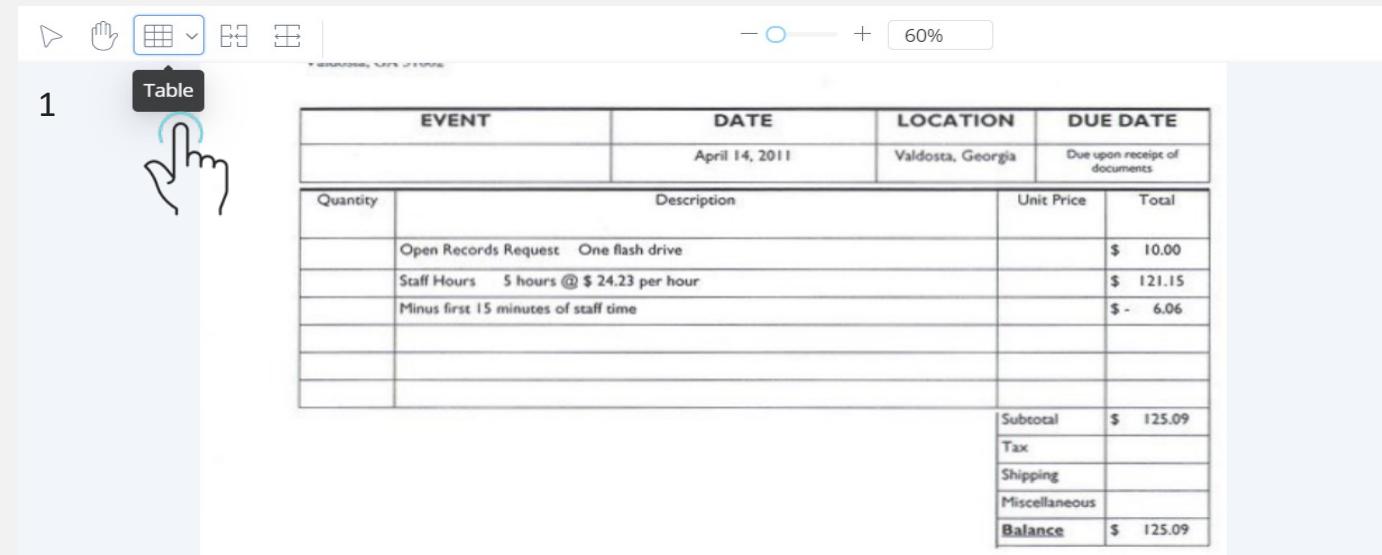


>> After going to the "Table" action menu the user can apply additional buttons "Merge Cells" and "Split Cells". The user can return to the «Label» menu.

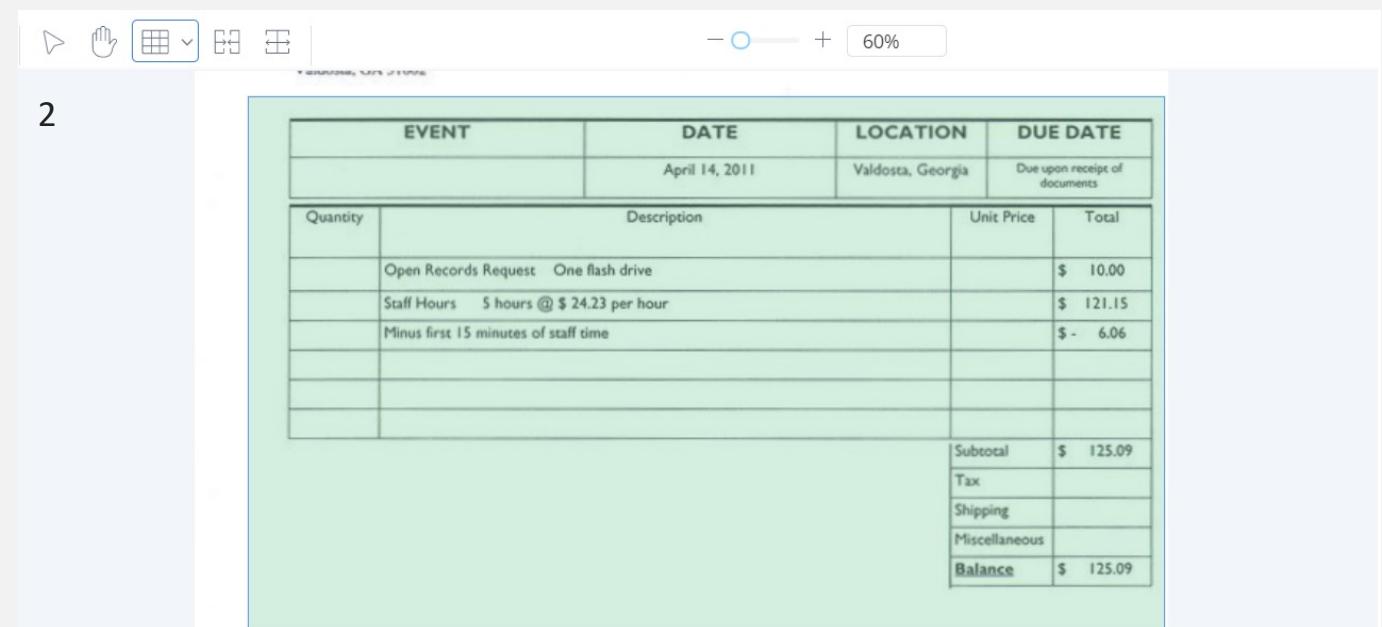
Manual table markup

The user can markup a table manually.

Step 1. The user opens the document in the Labeling Tool and must switch from the "Label" menu to "Table" and activate the table markup functions.



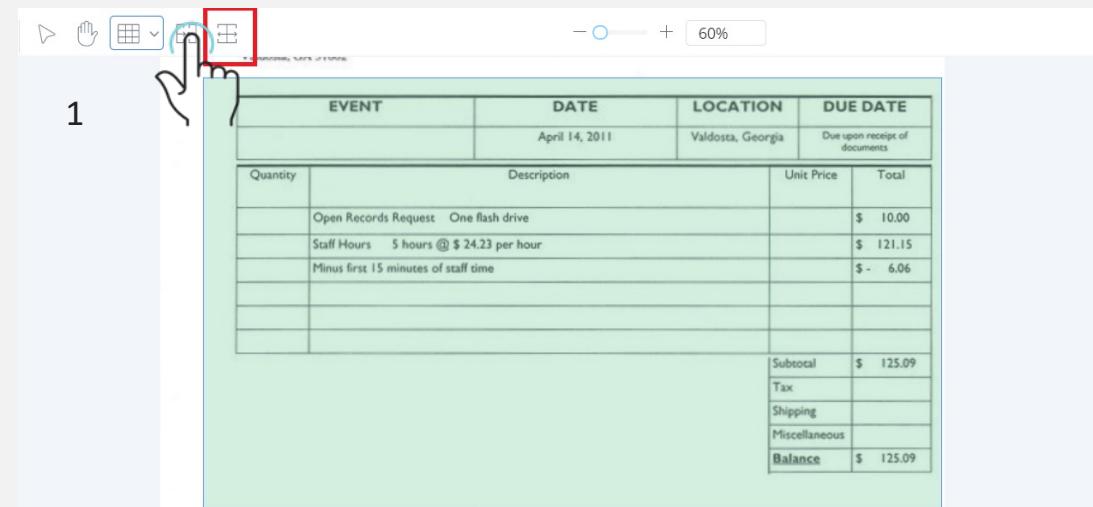
Step 2. By clicking the left mouse button, the user selects a table area in the document. The system highlights the selected area in green. The user can now split the table into rows and columns to form a suitable table markup. This can be done using the «Split Cells» and «Merge Cells» buttons.



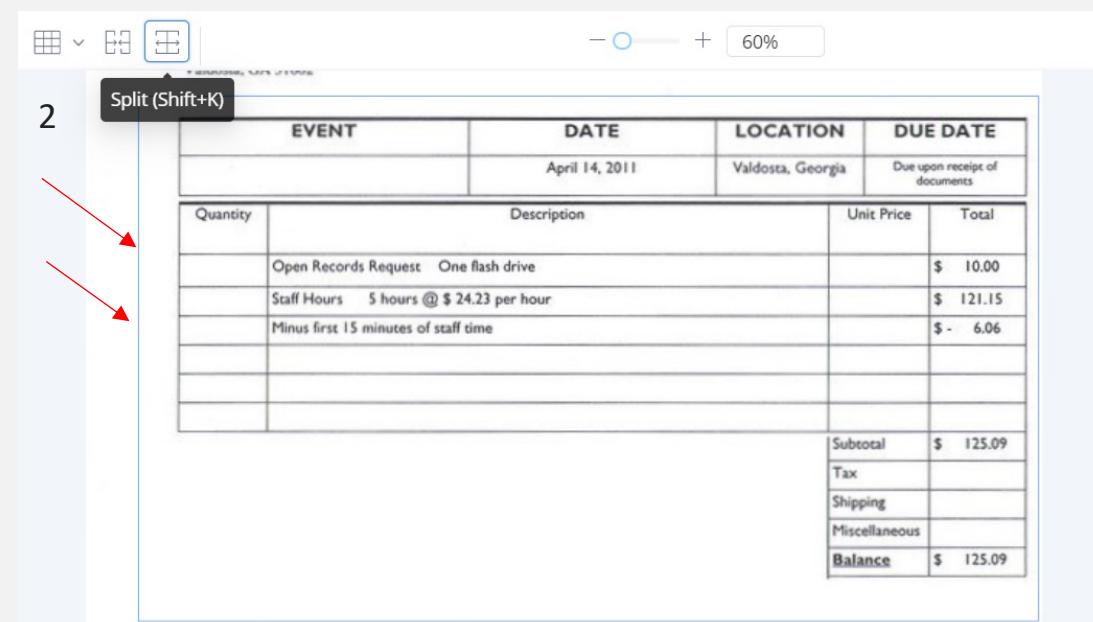
Split table cells

Step 1. Before splitting cells, the user must select a table area in the document.

User clicks the “Split Cells” button (Shift + K).

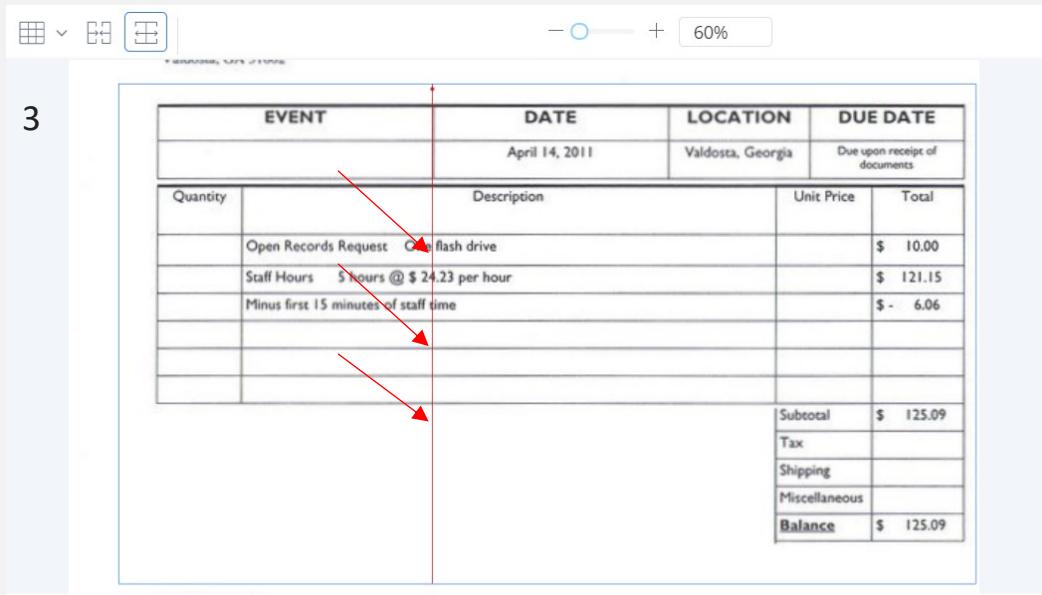


Step 2. When the user clicks the “Split Cells” button, the green selection of the table area disappears, leaving only the blue outline. The user can start drawing columns and rows.

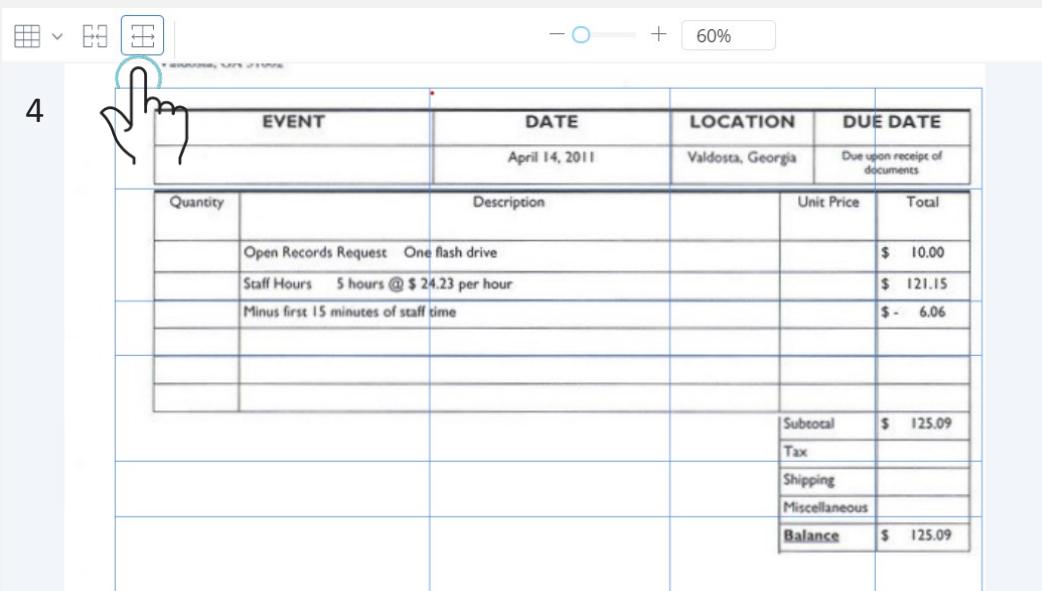


Split table cells

Step 3. By placing the mouse cursor in the desired location, the user begins to draw a column in the table area while holding down the left mouse button. During the drawing process, a red line appears, which turns blue when completed. When you press and hold the «Ctrl» button, the rows and columns are drawn automatically after the user specifies the exact location.



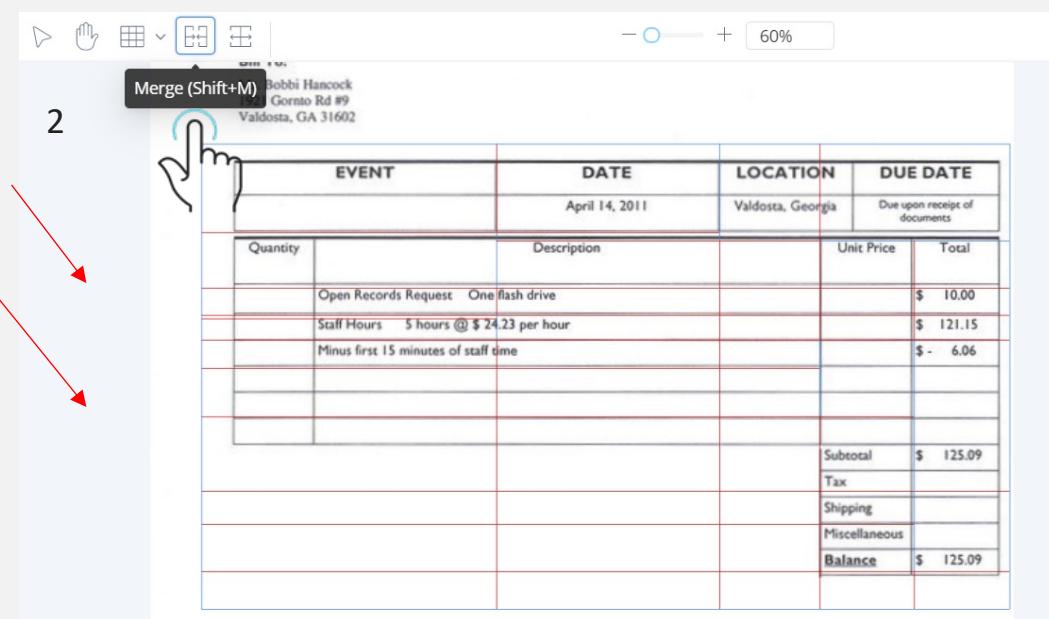
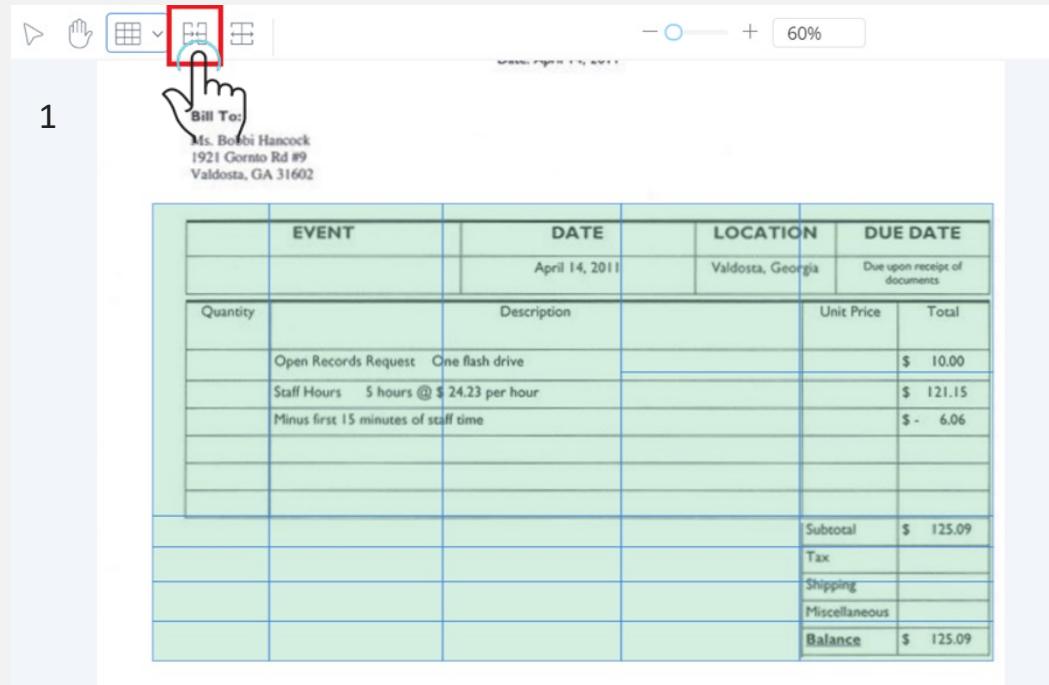
Step 4. Blue lines show user-drawn columns and rows using the «Split cells» function.



Merge table cells

Step 1. Before merging cells, the user must select a table area in the document. User clicks the «Merge Cells» button (Shift + M).

Step 2. When the user clicks the «Merge Cells» button, the green selection of the table area disappears, leaving only the blue outline and red markup inside the table. The user can start choosing the area to merge.



Merge table cells

Step 3. By placing the mouse cursor in the desired location, the user begins to select table cells for merging by holding down the left mouse button. As the user selects, cells are highlighted in red, and a blue rectangle appears in the selection area to indicate which area will be merged.

3

The screenshot shows a Microsoft Word document with a table. The table has columns labeled 'EVENT', 'DATE', 'LOCATION', and 'DUE DATE'. The 'DATE' column contains 'April 14, 2011'. The 'LOCATION' column contains 'Valdosta, Georgia'. The 'DUE DATE' column contains 'Due upon receipt of documents'. Below the table is a table of contents. A red arrow points from the text above to the first row of the table, and another red arrow points to a cell in the table of contents that is highlighted with a blue rectangle.

EVENT	DATE	LOCATION	DUE DATE
	April 14, 2011	Valdosta, Georgia	Due upon receipt of documents
Quantity	Description	Unit Price	Total
Open Records Request	One flash drive	\$ 10.00	
Staff Hours	5 hours @ \$ 24.23 per hour	\$ 121.15	
	Minus first 15 minutes of staff time	\$ - 6.06	
		Subtotal	\$ 125.09
		Tax	
		Shipping	
		Miscellaneous	
		Balance	\$ 125.09

Step 4. The selected cells are merged.

4

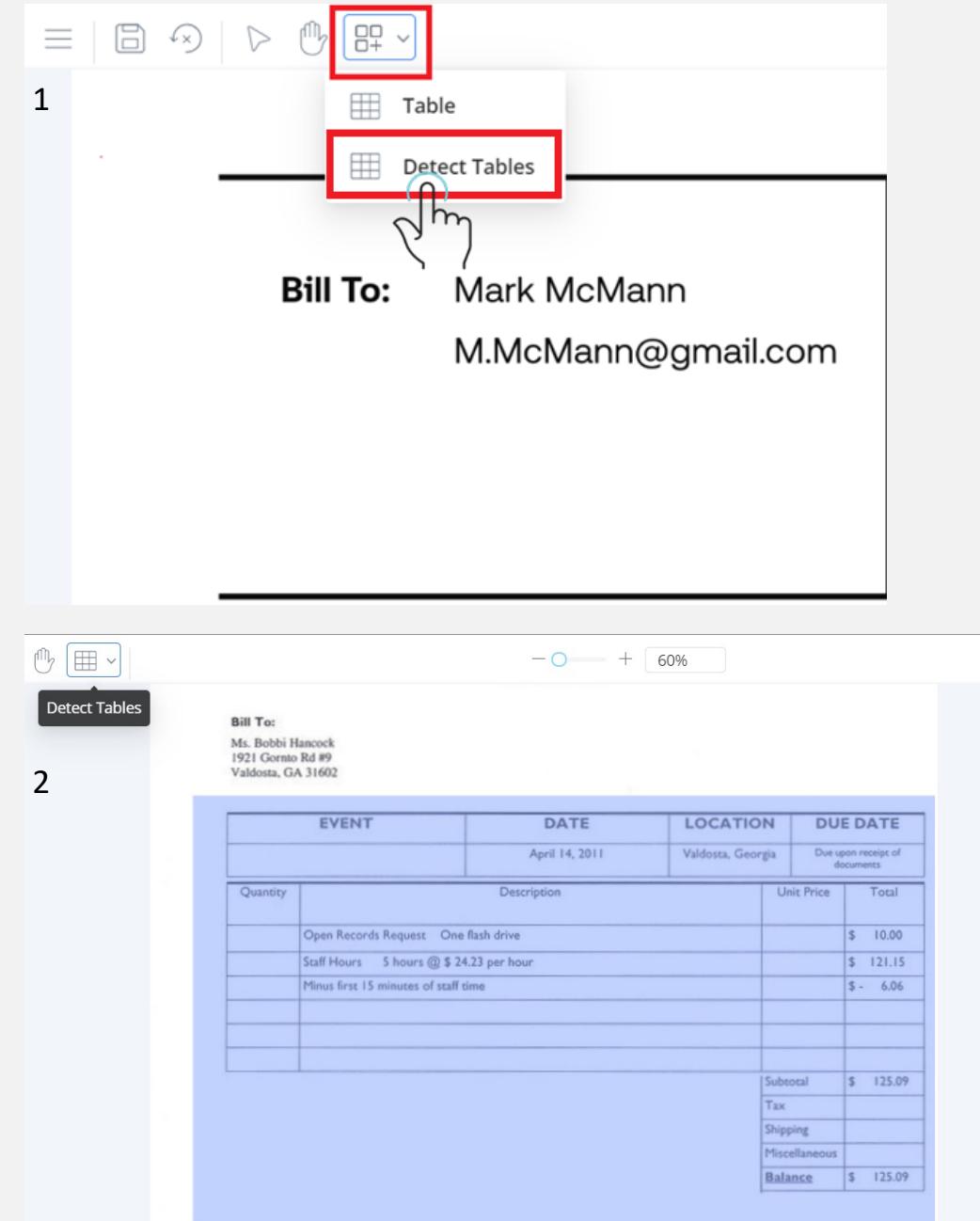
The screenshot shows the same Microsoft Word document after the merge operation. The table now has a single cell spanning the width of the first two columns. The table of contents remains unchanged. A red box highlights the merged cell in the table.

EVENT	DATE	LOCATION	DUE DATE
	April 14, 2011	Valdosta, Georgia	Due upon receipt of documents
Quantity	Description	Unit Price	Total
Open Records Request	One flash drive	\$ 10.00	
Staff Hours	5 hours @ \$ 24.23 per hour	\$ 121.15	
	Minus first 15 minutes of staff time	\$ - 6.06	
		Subtotal	\$ 125.09
		Tax	
		Shipping	
		Miscellaneous	
		Balance	\$ 125.09

Detect tables

The user applies the Detect Tables feature for searching and labeling tables automatically.

Step 1: Before detecting tables, the user should switch from the «Label» menu to «Table». User clicks the «Detect Tables» button on the «Table» menu.

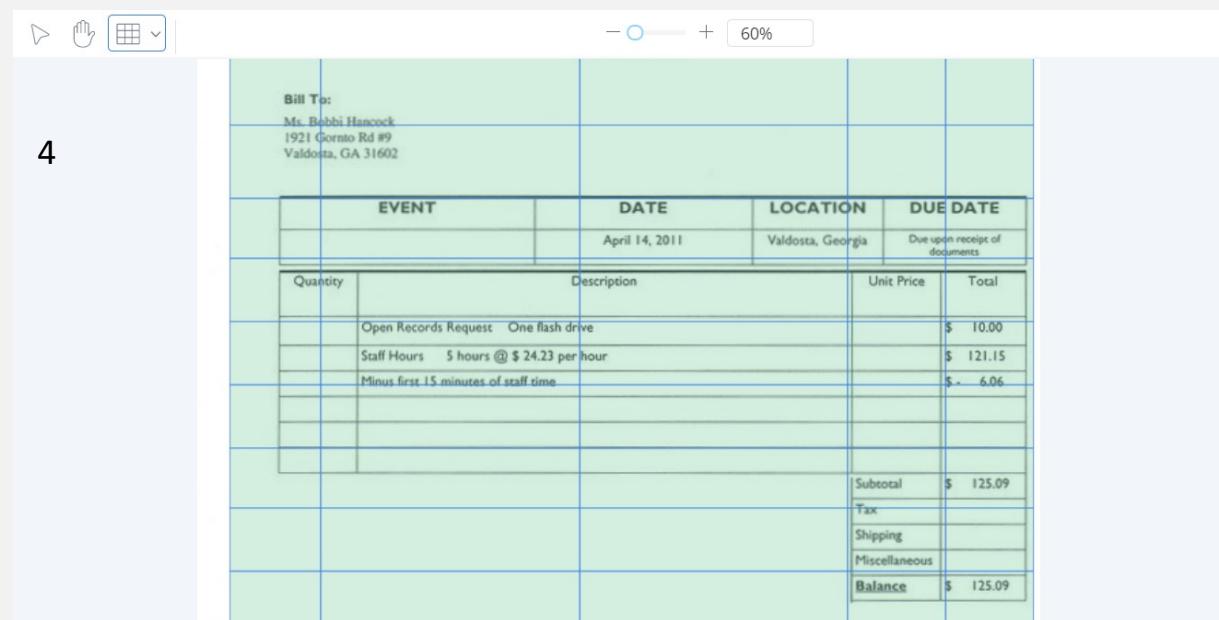
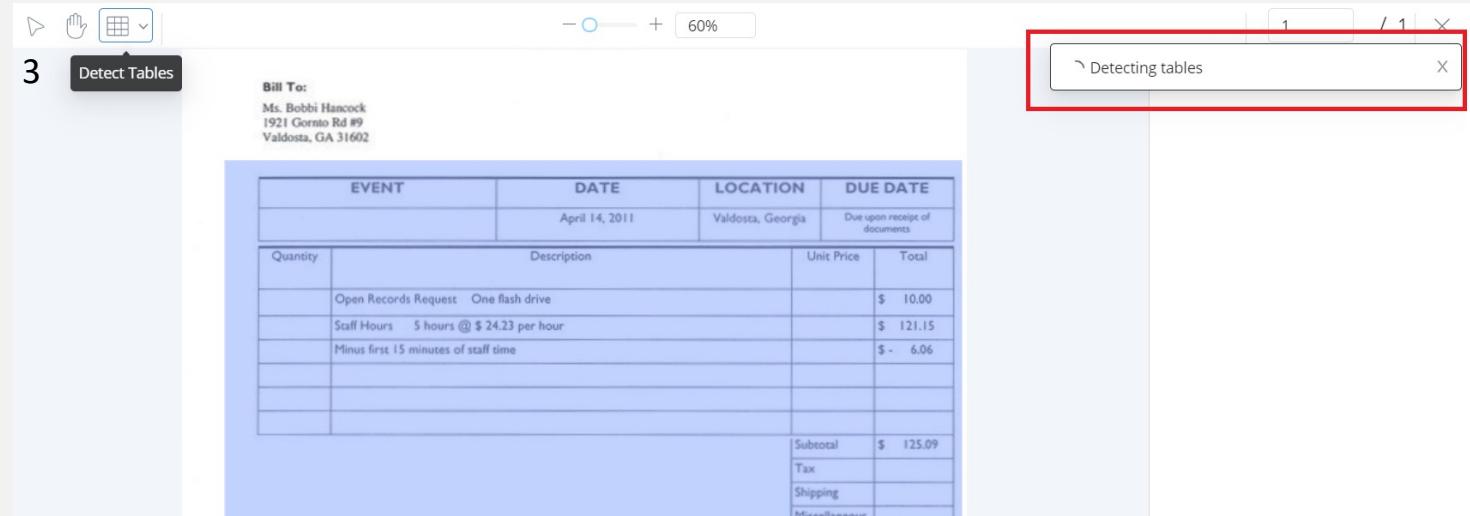


Detect tables

Step 3: After selecting the area to search for tables, the detection of tables in the document starts automatically. The user sees the info message "Detecting tables" and should wait for a little time.

When table detection is complete, a green area with blue markup appears in the highlighted area if the table is found.

The table found by the system does not always duplicate the markup of the table in the document. Requires manual adjustment of the table markup.

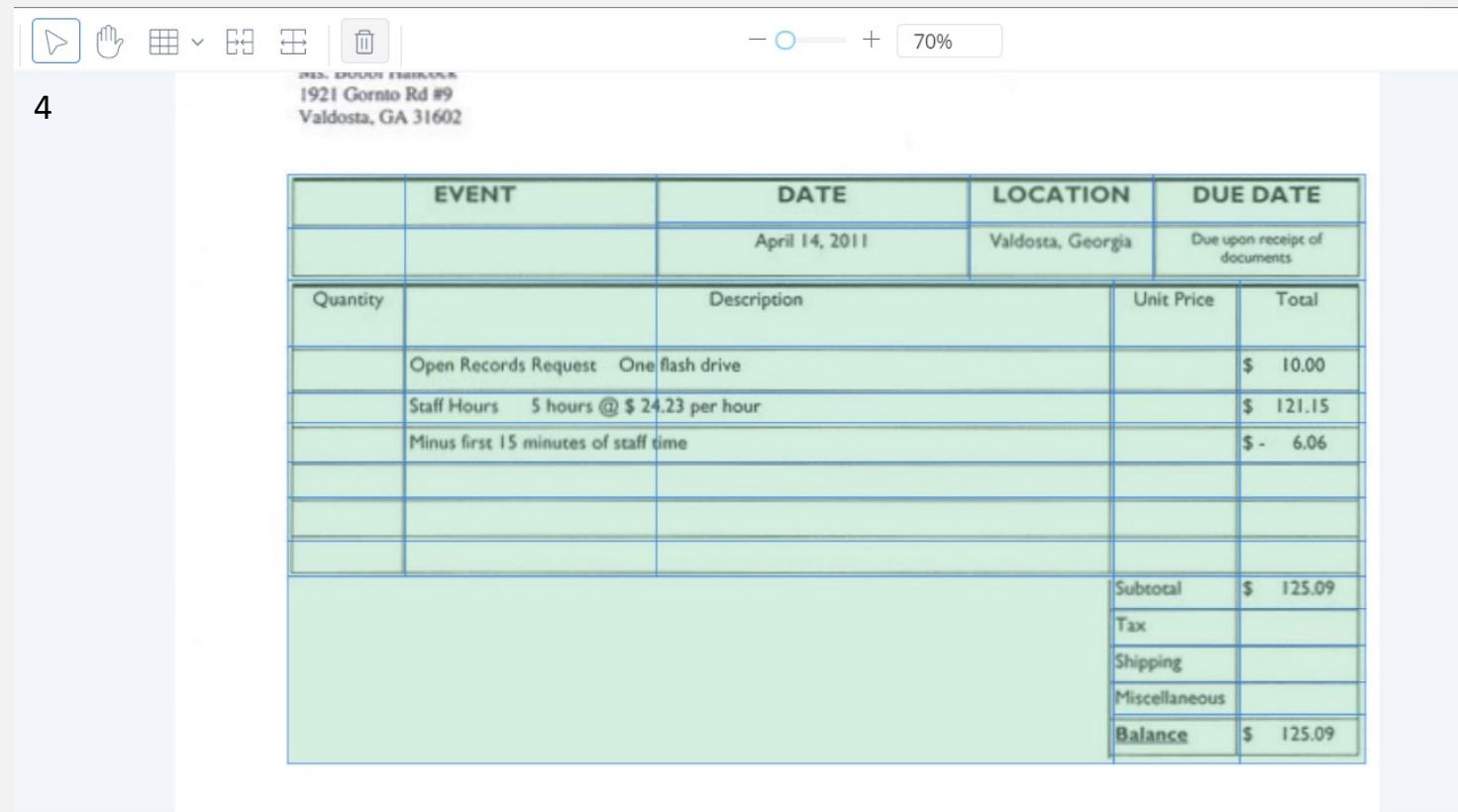


Detect tables

Step 4: The user can change the markup by moving it, splitting, and merging cells and columns (selecting the Pointer tool).

All actions are performed in order to make the markup in the Labeling Tool as similar as possible to the table in the document.

Right-clicking opens the table markup area editing panel. The user can add/delete row or column.



This is how the table detected by the system after editing looks like. To edit the table, the user applies splitting and merging of cells, as well as tools Pointer and Drag.

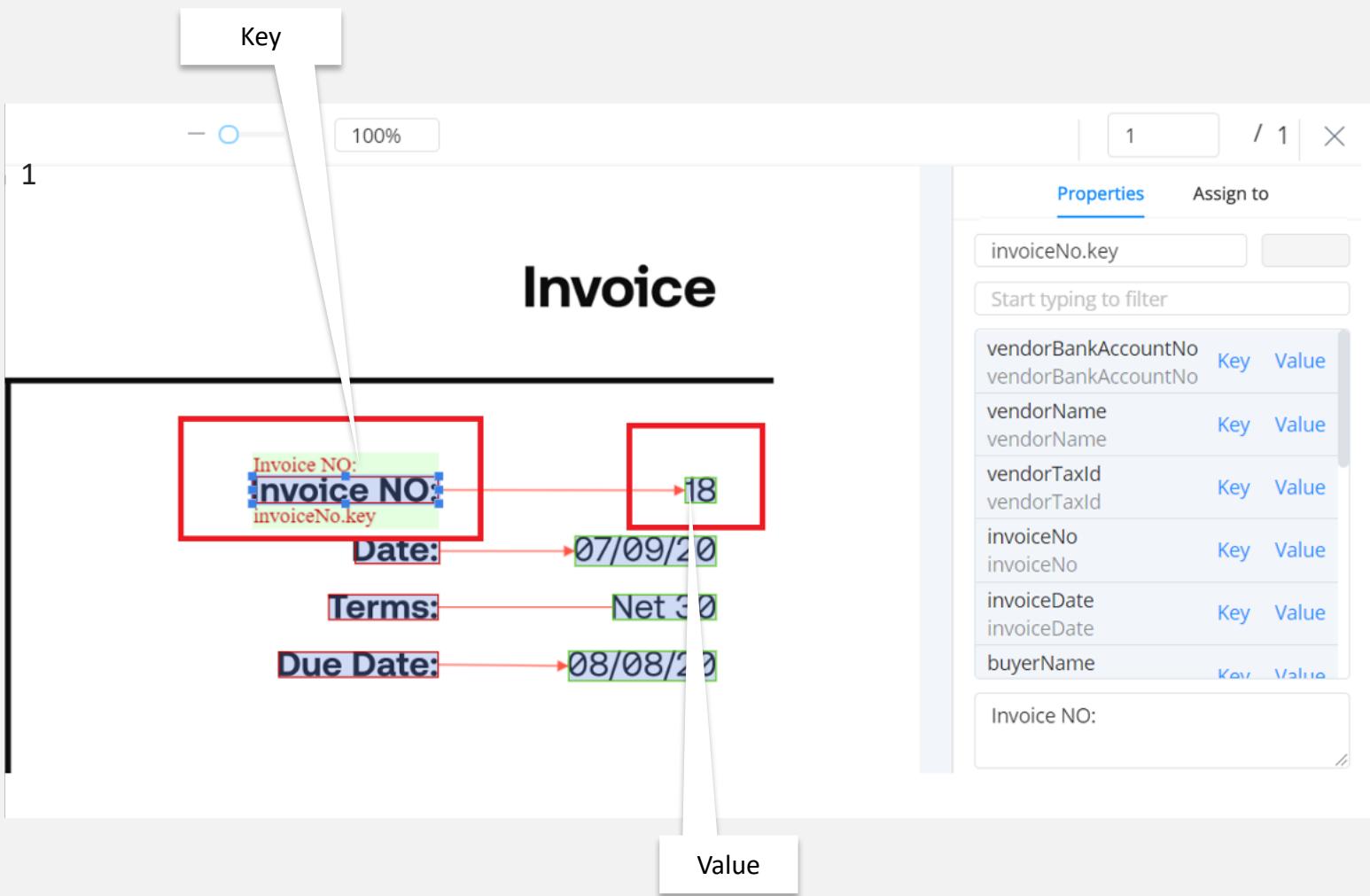
Key-value Pair Labeling

A key/value pair consists of two related data items: a key, which is a constant that defines a dataset, and a value, which is a variable that belongs to that dataset. A fully formed key-value pair might look like Figure 1.

When the user marks a key-value pair in a document, the flagged fields are marked as "Key" and "Value".

The user can label the key/value pair by assigning these properties in the sidebar to the right.

One key corresponds to one value, and these two fields are related to each other (for example, the "Account No." field and the account number "18").



Key-value pair Labeling

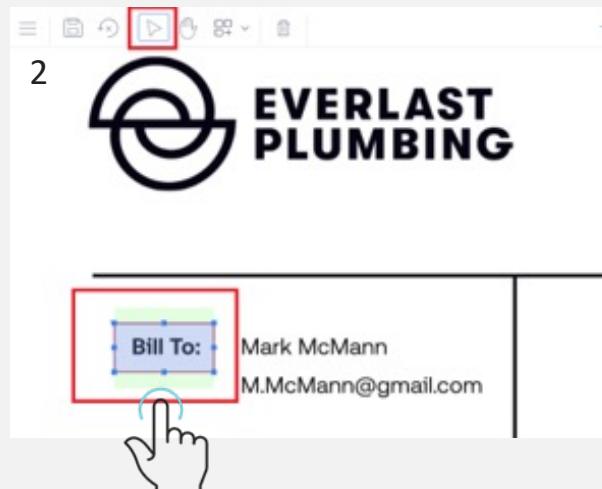
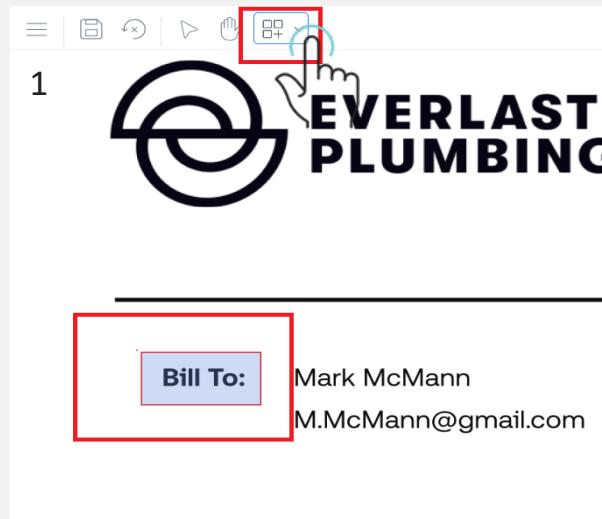
Step 1. The user opens the document in the Labeling Tool and must switch to the Label menu to activate the labeling functions.

The user selects an area for markup and draws a label to be a “Key”.

The selected area is highlighted in blue with red edges.

Step 2. The user selects the «Pointer» tool and selects highlighted area.

The selected area is highlighted in green color.



Key-value pair Labeling

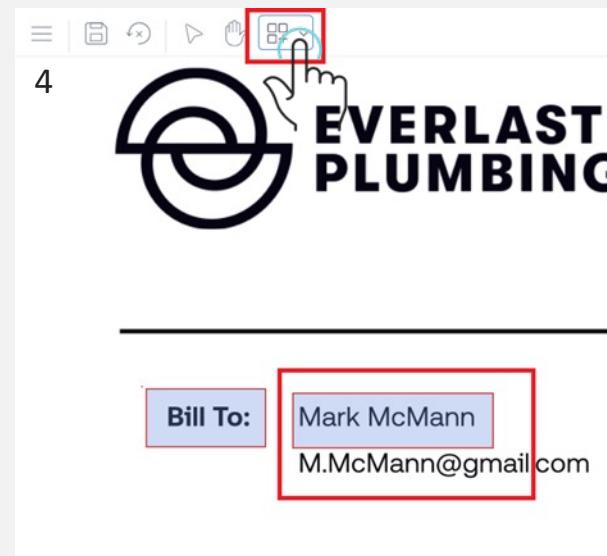
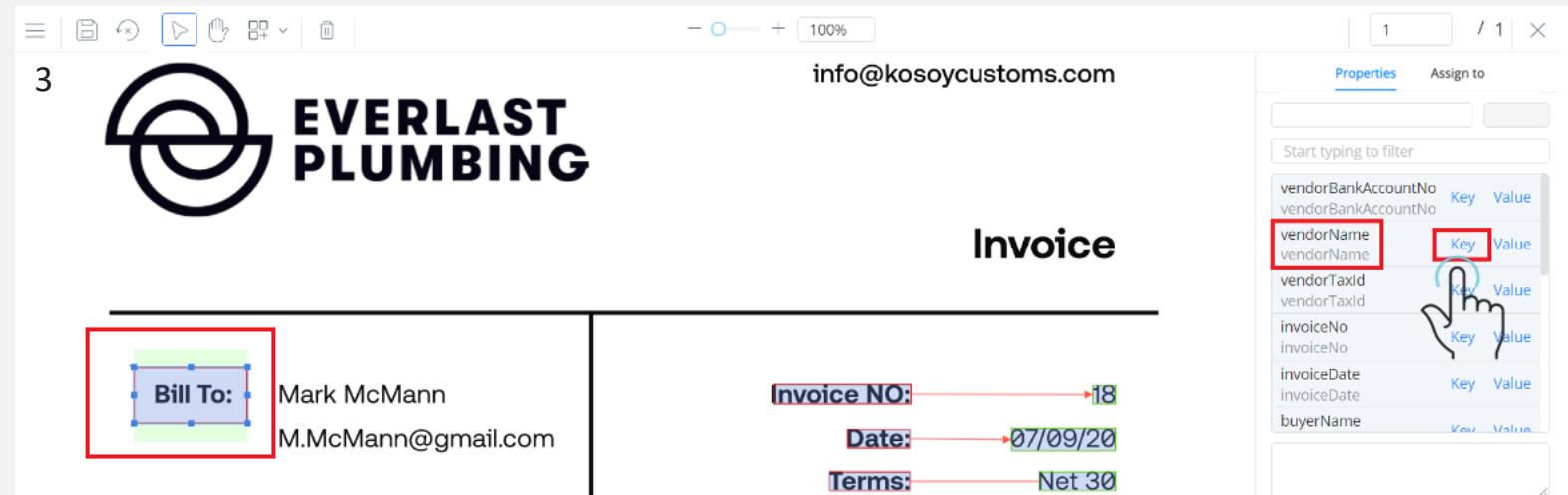
Step 3. To assign a value to a «Key» field, the user clicks on the word «Key» next to the name of the selected field.

The selected field is assigned the value «Key» .

Step 4. The user chooses «Label» to activate the labeling functions.

The user selects an area for markup and draws a label to be a «Value».

The selected area is highlighted in blue with red edges.



Key-value pair Labeling

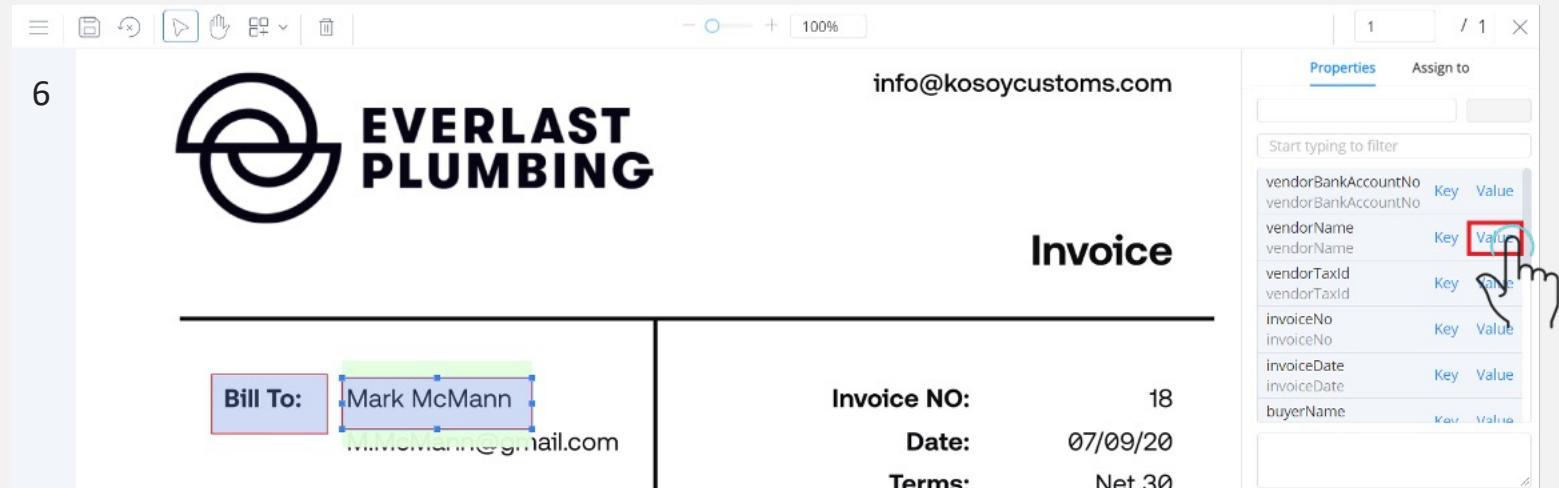
Step 5. The user selects the «Pointer» tool and selects highlighted area.

The selected area is highlighted green color.



Step 6. To assign a value to a «Value» field, the user clicks on the word «Value» next to the name of the selected field.

The selected field is assigned the value «Value».



Key-value pair Labeling

Step 7. The key-value pair was assigned successfully (key and value are related now).

The «Vendor Name» field now has a key and a value.

The key field has a red outline, and the value field has a green outline. An arrow between the fields indicates that the values are related.

