

OHG Action Invoice Scanner Instructions

This tool is still in its first public stages, it currently works by searching text extracted from PDF files, looking for patterns and matching them to a list of known contractors and addresses.

The next iteration will use machine learning to do this more automatically with less input. However, for now, I think it's still very useful for people to scan large numbers of receipts in an hour or two rather than the days it would otherwise take

If you are so inclined, the source code is available to be inspected here:

https://github.com/ohgaction/service_charge_scanner/blob/main/source/OHG_Action_Receipt_Processor.py

Please follow the installation instructions for how to get started and launch the invoice scanner. When the scanner loads you will be met with a menu that looks like this:

```
=====
Invoice Scanner - 2022 - E.Spencer
Version 1.0 Beta

This invoice extraction tool is free for personal use.
It comes with no warranty and is used entirely
at your own discretion. By proceeding you accept these terms.

=====

[R] Restore from disk          [P] Process
[F] Select input folder       [Q] Quit

Input folder: /Users/edspece/Desktop/Scripts/service_charge_scanne
=====
█
```

You can now press the F key to select a folder of receipts to scan. A dialogue window will appear. Select where you've saved your PDF files.

Press the P key and it will start processing. Once it has finished it will display a text display of each receipt. You can navigate between the receipts using the < or > keys, and you can scroll up and down using [and] keys (*you may need to stretch the terminal window open a bit, as it sometimes crops the top of the display*)

```

=====
Invoice 1/11
=====
Press [ or ] to scroll up and down
Press < or > to move between invoices
=====
Extracted Text:
=====
One Direct Maintenance Limited

b
ONEDIRECT

PART OF ONEHOUSING GROUP

Suttons Wharf South Date: 04/03/2019
44 Palmers Road
London
E2 OTA Invoice number: J0KVC0
Date Job Completed: 04/03/2019
Bill To Prope
One Housing Group UPRN / UCARN: = PALMER550000
. , Communal Area, Cradford
100 Chalk Farm Road Site name : House South
London Post code: E20DG
NW1 8EH
Job Code Description Cost

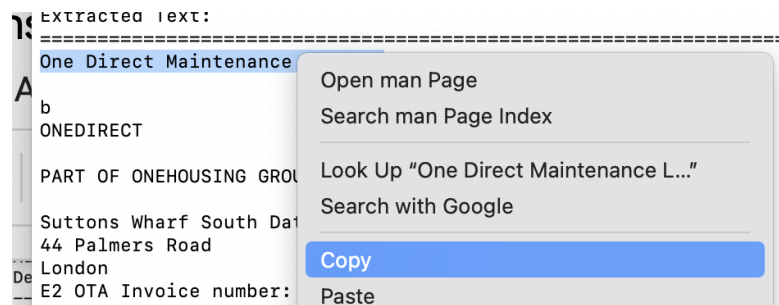
=====
File Name: J0KVC0.PDF
Date: 2019-03-04
Total: 218.54
Contractor: One Direct Maintenance Limited
Address: Suttons Wharf South
Description:

[S] Show original invoice file [A] Add contractor
[D] Add address [C] Change total
[T] Trim description

[X] Eport to CSV [Q] Save and Quit
=====

```

Below the scrolling part, which shows the text it has extracted from the invoice, you will see the fields it has collated. If the contractor is not displayed, I recommend copying the name of the contractor from the extracted text:



You can then press the A key to add the contractor to the database. You can paste (CMD+V or CONTROL+V) the contractor here.

```

=====
Please Enter Contractor: One Direct Maintenance Limited
=====

```

Once you hit enter, you should see that all receipts with this contractor have been updated when you navigate < and > through the receipts. If you find another receipt with a missing contractor, repeat until all the contractors have been filled in. You only need to add the names of

contractors once, after this they are stored in the database for each subsequent time you use the tool.

You essentially do the same to detect the address (if you want this in your data). But instead of using the A key, you use the D key to add an address to the database.

The total is automatically calculated by creating a list of totals from the receipt and then taking the highest value. This normally works fine, but you may find a few receipts where this is inaccurate. If you want to change a total press the C key.

You can use the S key at any time to show you the original PDF file

The Trim description can be used if you press the T key. You can now use the [and the] keys to trim down the receipt data to only show the description. Press T again to save this. In the next version, I'm going to look at doing this automatically using machine learning to detect descriptions.

When you are finished checking the invoices, press the X key and it will export a CSV file to the script folder. If you have installed as per the installation instructions this folder will be on your desktop called with Invoice_scanner_Windows or Invoice_scanner_macOS

output

J0KVC0.PDF	2019-03-04	218.54	One Direct Maintenance Limited	Suttons Wharf South
J0KYGU.PDF	2019-02-22	506.88	One Direct Maintenance Limited	Suttons Wharf South
J0L72K.PDF	2019-03-15	73.05	One Direct Maintenance Limited	Suttons Wharf South
J0KWER.PDF	2019-02-27	65.09	One Direct Maintenance Limited	Suttons Wharf South
J0L72N.PDF	2019-03-26	194.77	One Direct Maintenance Limited	Suttons Wharf South
J0L2WL.PDF	2019-03-06	472.45	One Direct Maintenance Limited	Suttons Wharf South
J0L8X7.PDF	2019-03-13	154.22	One Direct Maintenance Limited	Suttons Wharf South
678553.pdf	2019-07-01	13860.0	Brit-Sec Limited	
J0L75P.PDF	2019-03-07	165.31	One Direct Maintenance Limited	Suttons Wharf South
J0KYVA.PDF	2019-02-28	35.7	One Direct Maintenance Limited	Suttons Wharf South
J0KXHQ.PDF	2019-02-20	829.66	One Direct Maintenance Limited	Suttons Wharf South

if you load this you should see your results as a spreadsheet. This file can be easily imported into Excel or Google Sheets:

<https://support.microsoft.com/en-us/office/import-or-export-text-txt-or-csv-files-5250ac4c-663c-47ce-937b-339e391393ba>

In the final version, it will automatically export Excel files and will allow you to change the output directory and filename.