Potato Audit

Get Started

In Command
cd to target directory and run *npm start*

```
C:\Users\shiyaom>cd Desktop/Shiyao

C:\Users\shiyaom\Desktop\Shiyao>cd potato-audit

C:\Users\shiyaom\Desktop\Shiyao\potato-audit>npm start
```

cd to src and run *python app.py*

```
C:\Users\shiyaom>cd Desktop/Shiyao/potato-audit

C:\Users\shiyaom\Desktop\Shiyao\potato-audit>cd src

C:\Users\shiyaom\Desktop\Shiyao\potato-audit\src>python app.py
```
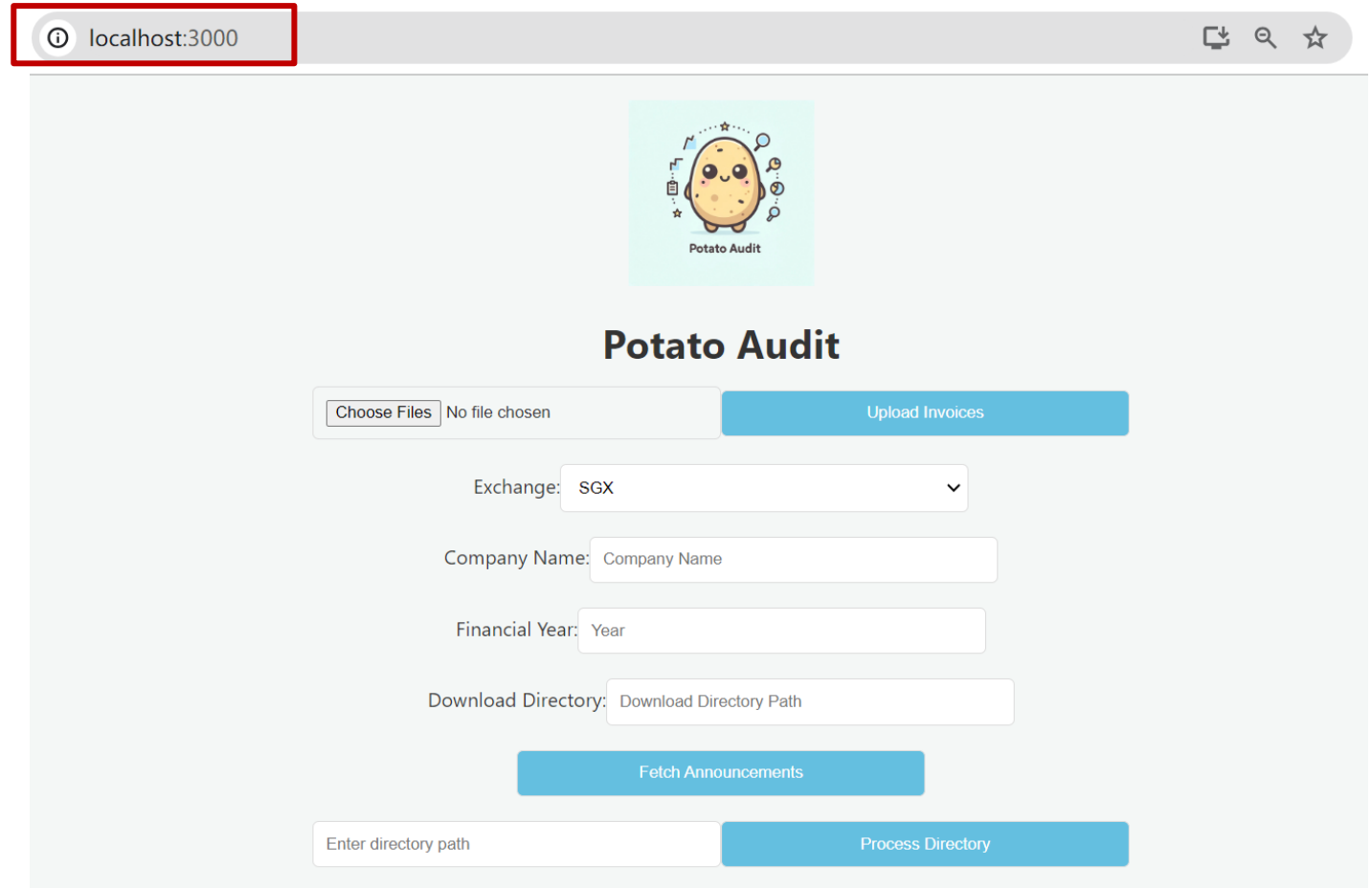
To Do:
Build package for the app so it could be started with a click

Potato Audit

Get Started

In Browser (Chrome)
Web UI is running on localhost 3000

Potato Audit

To Extract Data from Invoice/Structured Documents

Go to invoice folder

om\Desktop\Shiyao\potato-audit\Invoice_Test_Samples\TEST_ONLY_LARGE

Search TEST_ONLY_LARGE

Name	Date modified	Type	Siz
INV1298	23/Mar/2024 3:18 AM	Adobe Acrobat Docu...	

localhost:3000

Potato Audit

Choose Files | No file chosen | Upload Invoices

Exchange: SGX

Company Name: Company Name

Financial Year: Year
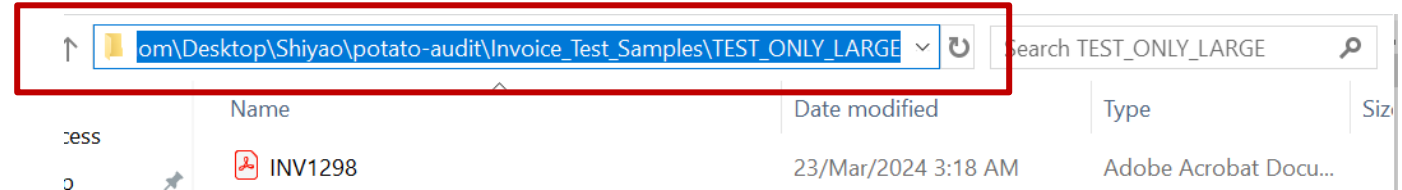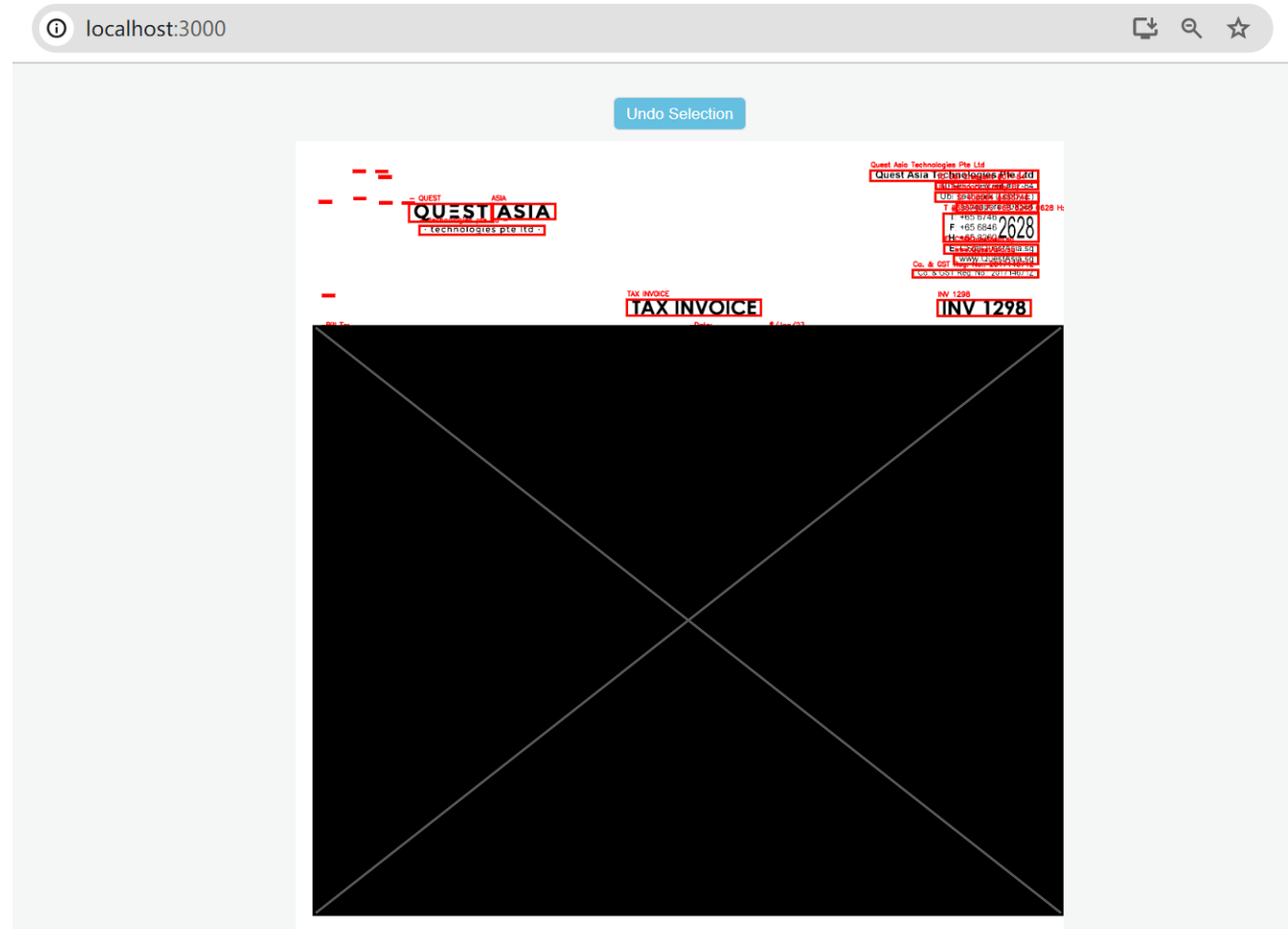
Download Directory: Download Directory Path

Fetch Announcements

Copy and paste directory path to input field

o-audit\Invoice_Test_Samples\TEST_ONLY_LARGE | Process Directory

Potato Audit

To Extract Data from Invoice/Structured Documents

Scroll down to find invoice image with annotated *bounding boxes*

Potato Audit

To Extract Data from Invoice/Structured Documents

**Enlarge the page if needed
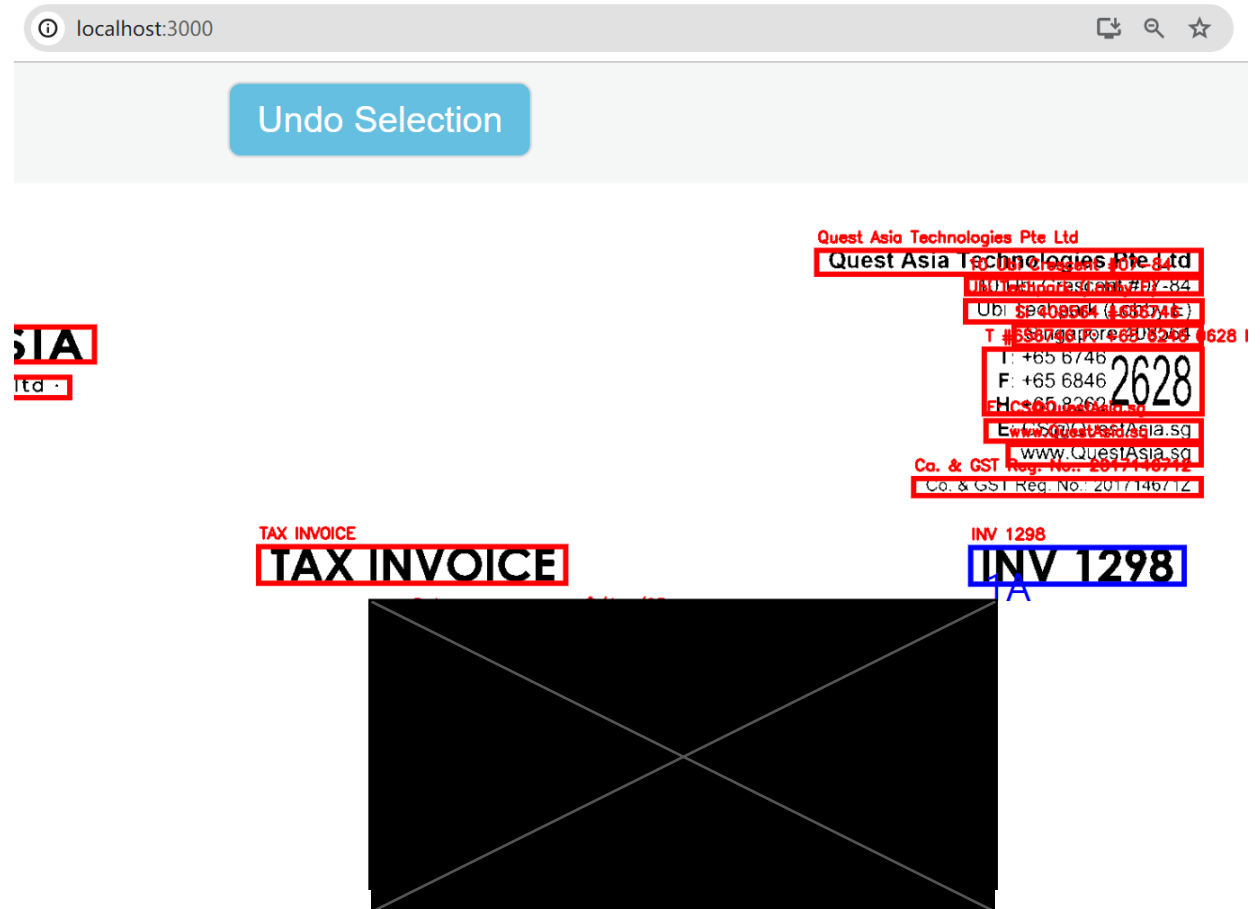Click on boxes to select *Field Label* (A) and *Value* (B) pairs that we need**

If a *Field Label* is also a *Value* itself, click on it twice

**Example:
Field Label: INV
Value: 1298
They are in the same box, so click on it twice**

localhost:3000

Undo Selection

Quest Asia Technologies Pte Ltd
Quest Asia Technologies Pte Ltd
Ubi
T
T: +65 6746
F: +65 6846
HC
E
www.QuestAsia.sg
Ca. & GST Reg. No.: 2017146712

SIA
Ltd

2628

TAX INVOICE
TAX INVOICE

INV 1298
INV 1298

Potato Audit

To Extract Data from Invoice/Structured Documents

Once a pair is selected, enter one or more *Field Label* variations, if any, separated by comma

Example:
Field Label: Bill To
Possible variations: Sold To

Potato Audit

To Extract Data from Invoice/Structured Documents

Click *Save* if there is any variation; else, *No Need*

Potato Audit

To Extract Data from Invoice/Structured Documents

If mistakenly selected a box, simply click *Undo Selection* to revert the change
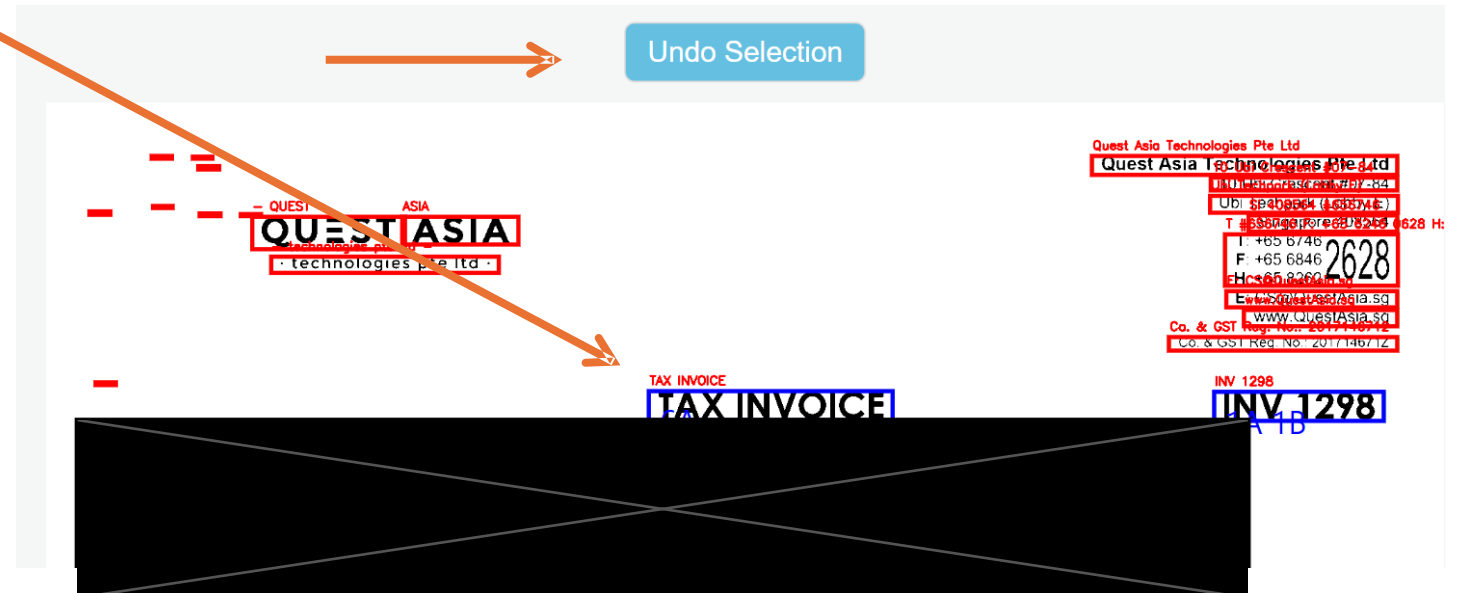
Undo Selection

Potato Audit

To Extract Data from Invoice/Structured Documents

Upon selecting pairs needed, click *Process Selection*
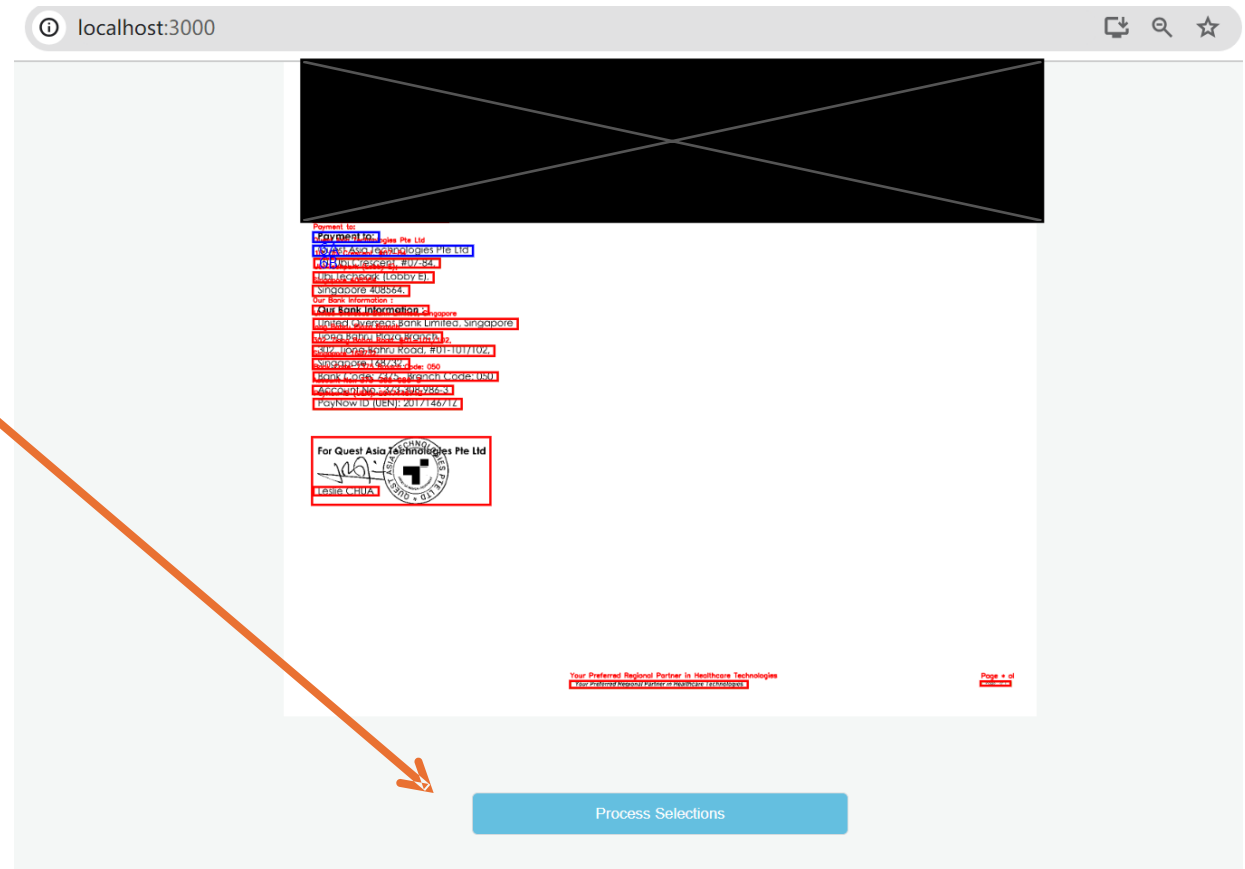
Potato Audit

To Extract Data from Invoice/Structured Documents

It is processing all the files in the specified directory, assuming they have similar format

Progress bar appears to track the status
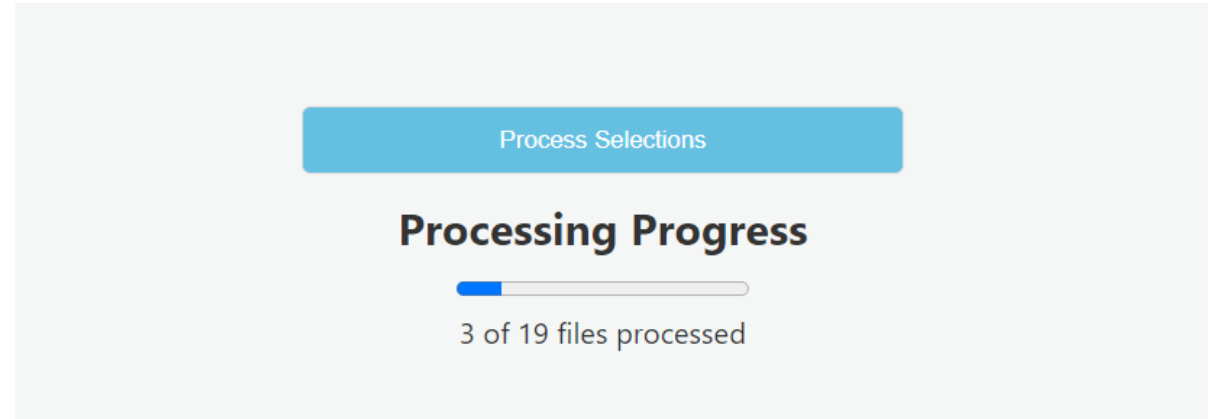
Potato Audit

To Extract Data from Invoice/Structured Documents

Once completed, a table appears below with sorting and search functions for preliminary inspection

Due to formatting issues, items may not be sorted correctly here

If there is missing or incorrect entries, copy the *File URL* in the last column to File Explorer to check the original file

Potato Audit

To Extract Data from Invoice/Structured Documents

For better post-processing, an Excel file with extracted data is saved to the same invoice directory named 'processed_invoices'
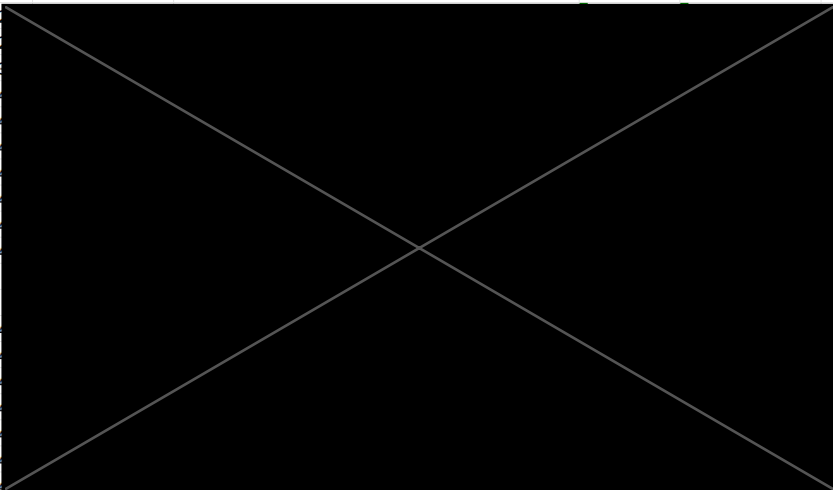
processed_info	3/29/2024 1:36 AM	JSON 源文件	8 KB
processed_invoices	3/29/2024 1:36 AM	Microsoft Excel 工...	7 KB

Potato Audit

To Extract Data from Invoice/Structured Documents

Open the Excel file to further clean and process extracted data

Potato Audit

To Extract Data from Invoice/Structured Documents

If there is any missing or incorrect data, simply click on the cell of the last column in given row, which directly pulls up the original file

Look for the information in the original file

Potato Audit

Performance Metrics

Training dataset of 175 sales invoices

					sia Technologies Pte Ltd	Click Here to Open File
177	Invoice Number	Buyer	Date	Sub-Total	Grand Total	Seller
178	158		175	170	140	173	173
179	90%		100%	97%	80%	99%	99%
180			Adjusted for blank sub-totals ->	173		
181				99%		

Processing time for training set ~350s
Processing time per invoice ~ <2s

To Do:
Test with invoices of other formats/languages

Potato Audit

Potential Use Cases with Invoice Data

Vouching and Reconciliation

100% Coverage of Samples

Visualise spending or sales patterns - check for anomaly (unusual volume/amount/date of transaction, duplicate payments, round amount invoices, or payments to unfamiliar vendors)

Potato Audit

To Fetch Announcements for Listed Co.

Select the exchange of the company (supports SGX/ASX; may integrate more exchanges if needed)

Enter the name/ID as recorded on the exchange (may consider doing a fuzzy match in the future)

Enter the financial year required

Specify a directory for the files to be downloaded to

Hit *Fetch Announcements*

Potato Audit

To Fetch Announcements for Listed Co.

Folder of <name of company>_<financial year> created at specified directory

Auto renaming of files with datetime prepended for easy sorting

Wait for the downloads

CSV file with extracted content created named 'Extracted_Announcements'

Potato Audit

To Fetch Announcements for Listed Co.

Require some manual reformatting – could be enhanced by passing the content to LLM for a summary if needed

Potato Audit

Performance Metrics

Downloading time for 50+ files (or any number of files for an entire year)
~ 2-3min max depending on Internet speed

Text extraction to CSV ~ < 10s

Manual Reformatting ~ < 10min

Note:
Texts extracted are the same as the announcements
Each announcement occupies one row (may consider using VBA to reformat things)