Schematic AI Extractor

Angie Ng, See Tien;
Kiu, Siang Hui; Lee, Zhe Jin; Yen Ming, Siaw
Intel Corporation,
Cloud Based Remote Debug (CBRD),
Penang, Malaysia

TestConX中国 China

Virtual ▪ November 21-23, 2023

intel

1

Contents

- Challenges
- Solutions
- Workflow
- Efficiency
- Technique
- Demo Video
- Summary
- Future Plans

TestConX中国 China

Schematic AI Extractor

2023

2

2

Challenges

- Reviewer required to search and go through pages which is labor-intensive.
- Manually identify the component and its details leading to inaccuracies.
- Saving the information by copying and pasting one at a time is time-consuming.

Solutions

- A program to perform automation extraction, allows faster extraction and reduces the time required to process.
- Improve the accuracy by implement Natural Language Processing (NLP) AI model to replace the search and identification process .
- Increase efficiency by saving all the information in one go.

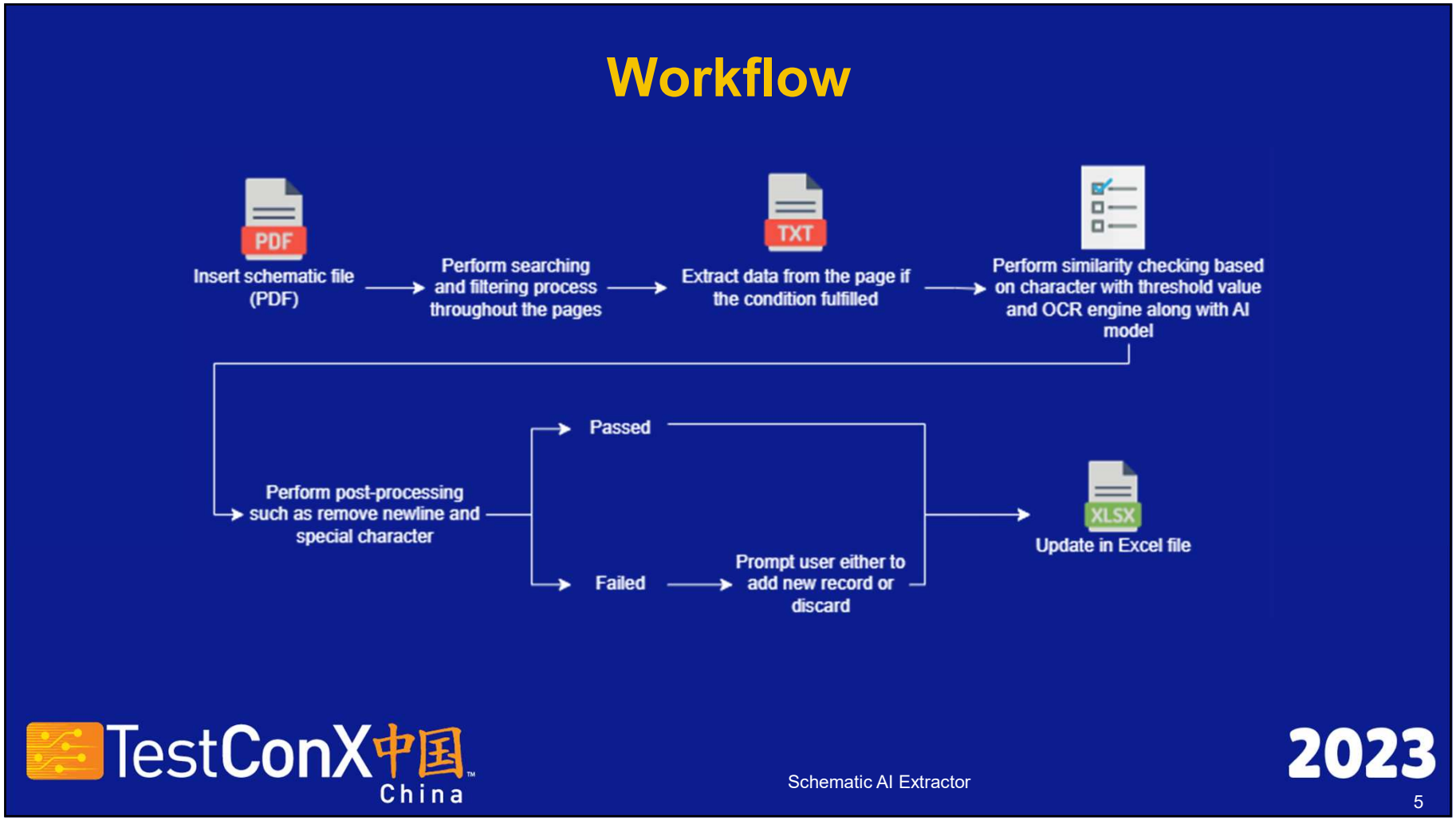
TestConX中国 China

Schematic AI Extractor

2023

4

4


Workflow
PDF
Insert schematic file (PDF)
Perform searching and filtering process throughout the pages
TXT
Extract data from the page if the condition fulfilled
Perform similarity checking based on character with threshold value and OCR engine along with AI model
Perform post-processing such as remove newline and special character
Passed
Failed
Prompt user either to add new record or discard
XLSX
Update in Excel file
TestConX中国 China
Schematic AI Extractor
2023
5

5

Efficiency

- Comparison on 30 schematic files

Time \ Method	Manual Extraction	Schematic AI Extractor
Processing Time (approx. hours)	5-6	1-2
Time Reduced (%)	60	

*extraction on one module

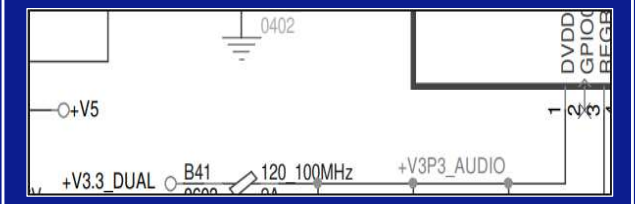
- Spelling error rate reduced as manual extraction for the reviewer might mistake on uppercase ‘i’ as lowercase ‘L’.

Technique

- Working with Python programming language, a sentence-transformers (all-MiniLM-L6-v2) model, PyMuPDF library, and Optical Character Recognition (OCR) engine to perform the extraction.
- The all-MiniLM-L6-v2 model aim is to be used as a sentence and short paragraph encoder. It used the pretrained (nreimers/MiniLM-L6-H384-uncased) model and fine-tuned with 1B sentence pairs dataset.
  - Small model size 80MB.
  - Fast processing speed.
  - Above average performance.

Technique

- PyMuPDF loads the schematic file and performs the search process.
- OCR engine recognizing the character patterns from image to extract into text.



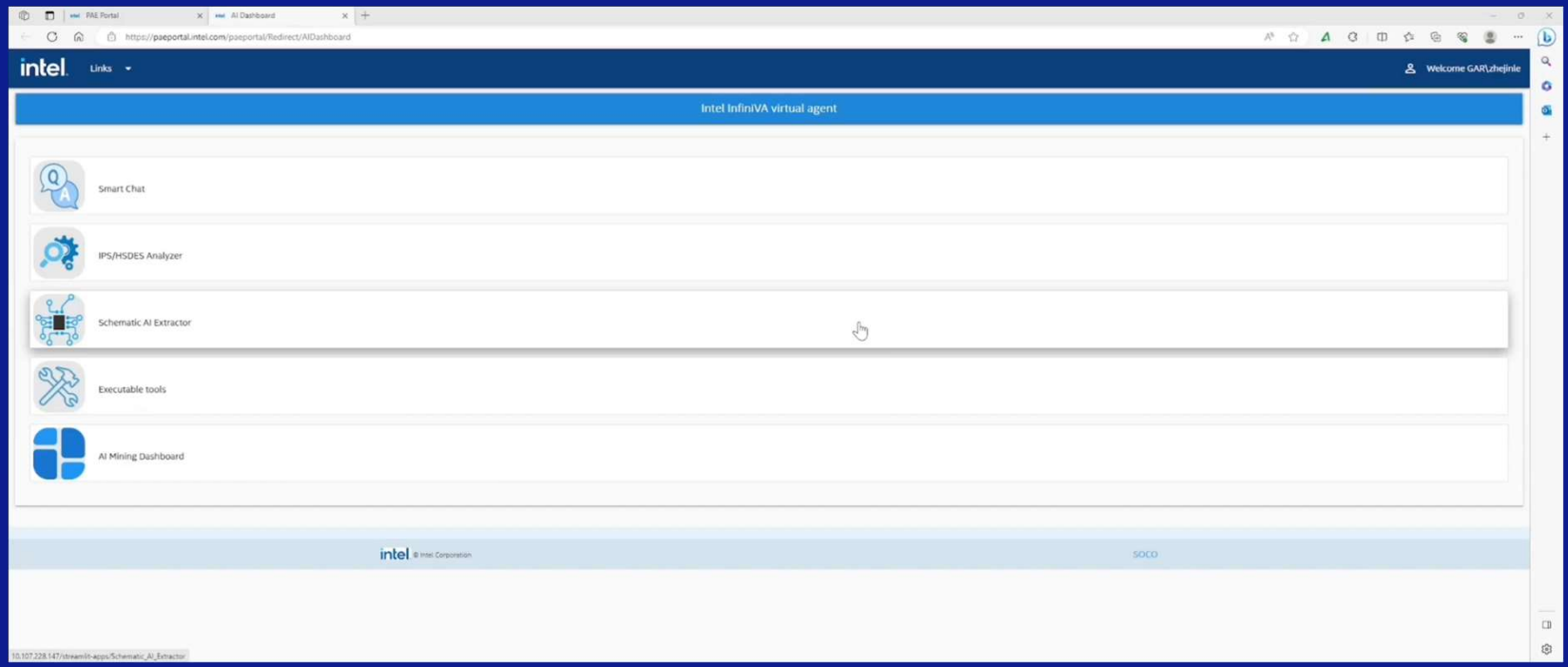
result.txt - Notepad

File Edit Format View Help

```
0402
DVDD
PIO
P
O+V5
23
B41
120
100MHz
+V3P3
AUDIO
+V3.3
DUAL
```

The post-processing result extracted from the image to text using the OCR engine

Demo



TestConX中国 China

Schematic AI Extractor

2023

9

9

Summary

- Schematic AI Extractor tool:
  - Reduces manual labor.
  - Improves extraction accuracy and processing time.
  - Other modules still work in progress.
- Current capabilities:
  - Automation extraction.
  - Saving extracted information automatically.
- Data usage:
  - Extracted data stored for future reference.
  - Enhances tool's performance.

Future Plans

- Able to perform automatic extraction with bulk schematic files.
- Add-on other module such as power, display, USB and PCIE.
- Possible to monetize this program to customer.
- Collaboration with other business unit or external customer to enhance this tool.

COPYRIGHT NOTICE

www.testconx.org

TestConX中国 China

2023

12