Version 1.0

2019/06/13

PDF ocr parsing using python

Development plan

Presented by: wang chao

iNeeds365

Dandong, liaoning, china

# PDF ocr parsing using python

The client has 1K pages of PDF and the contents is a list of lawyers.

The client needs to read this PDF file using OCR (Python) and parse the contents into their required format.

## Target Market

Potentially this module can be extended to future projects related PDF/OCR/TextParsing projects.

## prototype

N/A

## Roadmap

1. Reading PDF as images
2. OCR on images
3. Extracting valid list
4. Parse the content
5. Output to Excel

## timeline and milestone suggestion

| Task | Estimated Work Hours | Estimated Completion Date | Milestone |
| --- | --- | --- | --- |
| Reading PDF as images | 4 | 2019/6/3 | 120 |
| OCR module calibration | 16 | 2019/6/5 | 480 |
| Excluding not necessary data | 8 | 2019/6/6 | 240 |
| Parsing valid data | 8 | 2019/6/8 | 240 |
| Export to Excel file | 4 | 2019/6/9 | 120 |
| Test/Maintenance | 4 | 2019/6/10 | 120 |
| **TOTAL** | **44** | **2019/6/10** | **1320** |

### Budget

The total budget is 1176 and for further development requirements, I would charge 30USD/hr.

### Approval

| Title | Name | Date | Signature |
| --- | --- | --- | --- |
| Client |  |  |  |
| Developer | Wang Chao | 06/13/2019 |  |
|  |  |  |  |

# COmmunication

We will mainly communicate via Freelancer.com but we can use Skype, Whatsapp, Email after making a deal. Before making a deal, we are not allowed to communicate out of Freelancer.com