README.md 2023-10-08

Document AI (GCP) processing documents

Description:

The program (python source code) call the ocr and form parser Document AI (GCP).

The Document AI client API has interface in: C#, Go, Node, Python, Java, Php, Ruby

Document AI has two main processor:

• OCR processor:

You can use Enterprise Document OCR as part of Document AI to detect and extract text and layout information from various documents. With configurable features, you can tailor the system to meet specific document-processing requirements.

Bibliografy:

```
* https://cloud.google.com/document-ai/docs/enterprise-document-ocr
```

* https://cloud.google.com/document-ai/docs/process-documents-ocr

• Form Parser:

Form Parser v2.0 is a document processing model that can extract items from any document type, such as fields and tables. Combine it with Document AI's optical character recognition (OCR) to augment and automate extraction.

Example of Form Parser output with program source code:

```
************File Name form-cms1500.pdf****************
                                            Field Name
                                                              Field Value Confidence
0
                                  MEDICAID (Medicaid#)
                                    FECA BLKLUNG (10#)
                                                                            0.999752
                                    TRICARE (ID#/DoD#)
                                                                            0.999750
                                   CHAMPVA (Member ID#
                                                                             0.999737
                               GROUP HEALTH PLAN (10#)
                                                                             0.999035
                                                    YES
                                                                             0.998431
                                                    NO
                                                                             0.998064
                                                    YES
                                                                             0.997959
8
                                                    YES
                                                                             0.997585
                                                    NO
                                                                             0.995612
10
                                                    NO
                                                                             0.995199
11
                                           OTHER (ID#)
                                                                            0.994954
12
                                                    NO
                                                                            0.994008
13
                                                 Other
                                                                            0.991844
14
                                                                            0.991452
                                                 Spouse
15
                                                    YES
                                                                             0.987589
16
                                                  Child
                                                                             0.978465
17
                                  MEDICARE (Medicare#)
                                                                             0.970292
18
                                                                             0.962994
                                                  STATE
19
                         TELEPHONE (Include Area Code)
                                                                             0.960054
    2. PATIENT'S NAME (Last Name, First Name, Midd...
20
                                                                             0.906608
                      PATIENT'S ADDRESS (No., Street)
                                                                             0.872310
```

README.md 2023-10-08

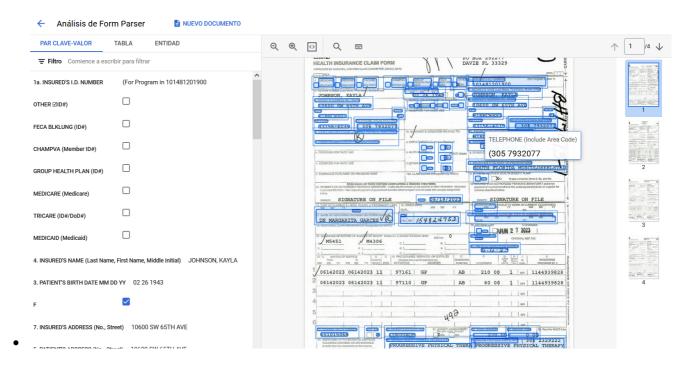
- * https://cloud.google.com/document-ai/docs/process-documents-form-parser
- * https://cloud.google.com/document-ai/docs/form-parser

• You can create or fine-tune other ocr document processor in Document AI if you have some examples for train and testing. There are a customize options in Document AI GCP platform.

Check tutorials:

- https://cloud.google.com/document-ai/docs/process-documents-client-libraries
- https://cloud.google.com/document-ai/docs/process-documents-form-parser
- https://cloud.google.com/document-ai/docs/process-documents-ocr

Example Form Parser (Document AI) processor:



Document AI (GCP) Billing

- Free on charge 1 million pages
- For each 1 million pages 1.5 dollars

Learning resources:

- * https://www.youtube.com/watch?v=6fylP4JMzjo
- * https://www.youtube.com/watch?v=F_jyoe1lQhg
- * https://medium.com/@rohit.gupta_47632/key-value-entity-extraction-from-forms-using-gcp-apis-9c59eb731dce
- * https://codelabs.developers.google.com/codelabs/docai-form-parser-v1-python#4