

# Arithmer OCR

## AI Systems



### OCR

Arithmer OCR converts printed, handwritten texts and image documents into machine-encoded texts by an advanced still image analysis system. Based on a highly advanced AI technology, we refined a recognition rate that goes beyond the accuracy of our competitors.

### Inspection

Arithmer OCR converts printed, handwritten texts and image documents into machine-encoded texts by an advanced still image analysis system. Based on a highly advanced AI technology, we refined a recognition rate that goes beyond the accuracy of our competitors.

### Dynamics

Based on a highly advanced AI video analysis, Arithmer Dynamics extracts data such as objects, locations, people and recognized human actions. The technology even reads subtle pixel changes that enables visualizing airflow in videos.

### NLP

Arithmer NLP provides an original Natural Language Processing (NLP) system that instantly extracts optimal answers from a large-scale corpus. The technology semi-automatically creates an AI Chatbot that conducts conversations from internal FAQ systems.

### DB

Arithmer DB automatically analyzes and links disorganized and dispersed big data through revising intelligent agent (IA) technology which is installed to an integrated database environment. Arithmer DB's IA autonomously analyzes and figures out organic data.

### R3

Arithmer R3 is a 3D modeling technology developed by the full appliance of advanced mathematics. Arithmer R3 reduces 3D Big data dimensions to key parameters, enabling 3D model building through utilizing human body and dental model measurements, and etc.

### Robo

Arithmer Robo provides high-tech robot systems converted from Arithmer, Inc.'s characteristic AI technology. Collaborating with the world's top-class robot manufactures, Arithmer Robo aspires to create AI smart robots beyond the skills of handwork.

## ArithmerOCR 紹介セミナー

Arithmer

2021/01/14

## 目次

1. OCRとは
2. OCRの出力形式
3. OCRの種類
4. 文字の種類
5. OCRのパッケージの紹介

# 1. OCRとは

- ・OCRは【Optical Character Recognition】の頭文字を取ったもの
- ・画像データのテキスト部分を認識し、テキストデータとして出力する

## 入力(画像データ)

請求書

アリスマー株式会社 御中 請求No. 123456789  
 ご担当: 有住真太郎 様 請求日 1234/12/12

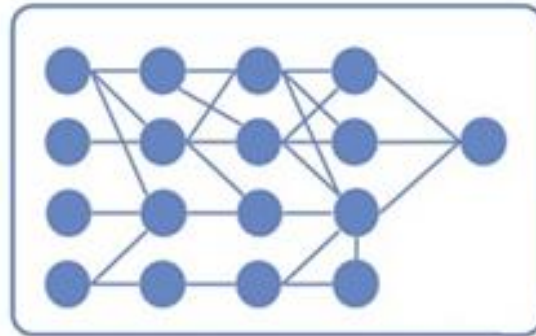
件名: AIシステム  
 下記の通り、ご請求申し上げます。

アリスマー株式会社  
 〒106-6040  
 東京都新宿区新宿1-3-3  
 新宿第1ビル2階  
 TEL: 03-5579-6683  
 FAX: 03-5579-6683  
 E-Mail: sample@arithmer.co.jp  
 担当: 担当者

合計金額 ¥1,100,000 (税込) お支払期限: 1234/12/12

No.	摘要	数量	単価	金額
1	AIシステム 式	1 式	1,000,000	¥1,000,000

## 機械学習により推論



## 出力(テキストデータ)

1. 請求先会社名  
アリスマー株式会社

2. 請求No.  
123456789

3. 担当  
有住真太郎

4. 請求日  
1234/12/12

5. 件名  
AIシステム

## 2. OCRの出力形式

・OCRの出力形式は大きく分けて2通りある

①出力したテキストと項目の対応はつけない形式

非定型帳票パッケージの出力形式

②出力したテキストと項目の対応をつける形式(キーバリュー形式。構造化されたデータともよんでいる)

定型帳票パッケージ、請求書パッケージ等の出力形式

顧客は②の形式を欲している場合が多い



OCR



Text1: 氏名  
Text2: キョウカイトロウ  
Text3: 協会太郎  
Text4: 生年月日  
Text5: 平成元年5月10日  
.  
.  
.

①の形式

氏名: 協会太郎  
フリガナ: キョウカイトロウ  
会社名: 平成元年5月10日  
電話番号: 03-5579-6683  
.  
.  
.

②の形式(キーバリュー形式)

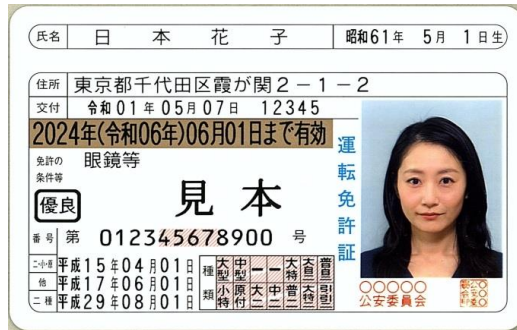
オレンジ色が項目名(キーとよぶ)

緑色が内容(バリューとよぶ)

### 3. OCRの種類

・OCRは大きく2種類に分かれる

①定型OCR・・・帳票のフォーマットが決まっているOCR(免許証、同じ会社の保険証券等)。OCRの難易度低め。フォーマットが決まっているため、読み取りたい項目の座標を指定することで、項目との紐づけが可能



②非定型OCR・・・帳票のフォーマットが決まっていないOCR (請求書、健康保険証等)。OCRの難易度高め。フォーマットが決まっていないため、読み取りたい項目の場所も決まっていない



## 4. 文字の種類

・文字の種類は大きく2種類に分かれる

①活字文字・・・プリンター等で印刷された文字。OCRの難易度低め。

活字例

A rectangular image showing the word "Arithmer" printed in a classic serif font on a light-colored background.

②手書き文字・・・人の手によって手書きで書かれた文字。OCRの難易度高め。

手書き例

A rectangular image showing the word "Arithmer" written in a cursive, handwritten style on a dark background.

## 5. OCRのパッケージ紹介

	非定型帳票パッケージ	定型帳票パッケージ	請求書パッケージ
対象帳票	定型でも非定型でも可能	定型帳票	請求書
出力形式	項目との紐づけは行わない	項目と紐づける	項目と紐づける
出力される文字	帳票中の全ての文字が出力される	座標指定された項目が出力される	過去のテキストデータの結果と、合計金額、日付が出力される
事前準備	事前準備の必要なし	読取りたい項目を事前に座標指定(枠を囲む)する必要あり	同じ会社の過去データ(画像+テキストデータ)を準備する必要あり
リリース時期	2019年度に顧客デモ向けに開発。製品化に向けて検討中。	昨年12月にテスト版リリース。	昨年9月にver1.0リリース。2か月おきにverアップ。
特徴	画像をアップロードしたら利用できるため簡単に使える。	類似画像検索等のInspectionの技術も用いている。まだ入れていない技術もある。	類似画像検索により過去のテキストデータの結果を利用することで、帳票に書かれていないデータを出力できる(帳票に書かれた金額を部署毎に振り分けた金額等)

人間に、愛を。  
未来に、AIを。

Arithmer 株式会社

〒106-6040

東京都港区六本木一丁目6番1号 泉ガーデンタワー 38/40F(受付)

03-5579-6683

<https://arithmer.co.jp/>

Arithmer

