

SmartRead 非定型文書強化バージョン SmartRead API V4 概要

2025/04/30

株式会社 Cogent Labs

本資料について



本資料は、「SmartRead 非定型文書強化バージョン」で使用可能な「SmartRead API V4」について解説しています。

特にバージョンの明記がない場合は「SmartRead API V4」の仕様解説になります。

V2, V3 については、V4との比較情報として一部掲載していますが、基本的には解説していません。

なお、本資料は既存資料である、APIドキュメント、マニュアルを補完する位置付けになります。必要に応じて各種ドキュメントも参照ください。

資料の内容

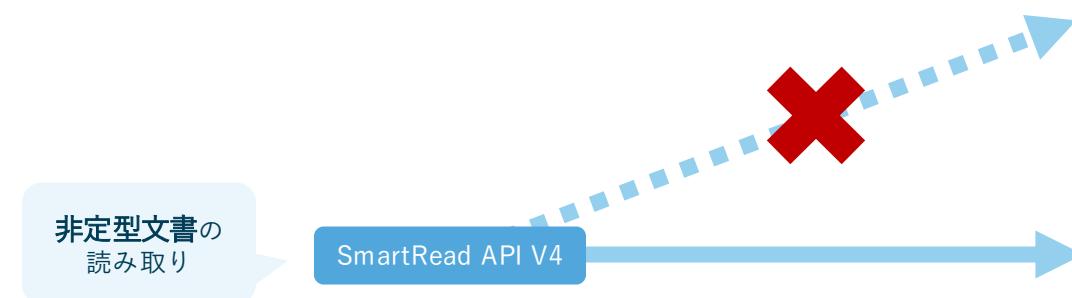
- 機能概要
- 目的別処理手順
 1. 読取結果の取得（エクスポート）
 2. ファイル画像データの取得
 3. 項目画像データ, 座標情報, 確信度 の取得
- 文書仕分け用API
- 参考

機能概要

SmartRead API のバージョンと使用可能なSmartReadについて



※ 一部の非定型文書の読み取り、テキスト付PDFファイルの作成が可能です。



COGENTDX
SmartRead

COGENTDX
SmartRead
非定型強化バージョン

※ 「SmartRead API V4」は、SmartRead 非定型強化バージョンに対してのみ利用可能です。

SmartRead API V4 で取り扱いが可能な読み取り対象文書の種類



読み取り対象文書/機能	対応状況	備考
非定型文書	○	2025年4月3日、自動仕分け機能追加リリース
定型文書	×	今後のリリースで対応予定
全文読み取り	×	今後のリリースで対応予定
表抽出 (SmartReadと同等機能)	×	今後のリリースで対応予定

結果の取得方法と種類

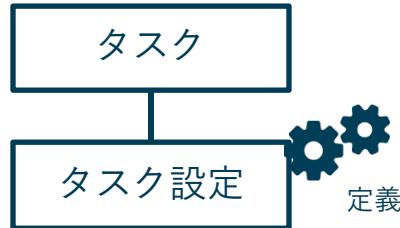
出力方法 (エンドポイント)	出力 形式	読み取 り項目/結果	座標情報	確信度	読み取り 項目画像URL	読み取り ページ画像URL
/v4/export/:exportId ※ WebUI同様、ZIPファイルでの取得になります。	CSV Excel JSON	○				
/v4/verification-data/:jobId/results/:pageId	JSON	○	○	○	○	
/v4/verification-data/images/:fileId	JSON					○

※ 画像は直接取得できません。取得用URLにアクセスすることで取得できます。なお、URLの有効期限は1時間です。

用語の解説

読み取処理概念の変更 (V3→V4) 1/2

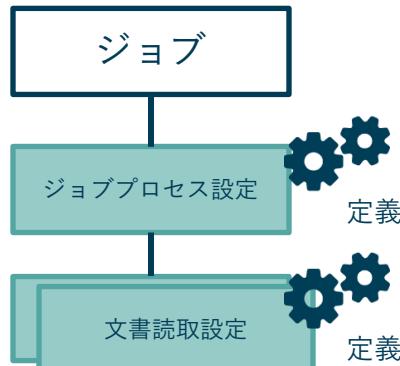
v3



- タスク/タスク設定

- タスク/タスク設定
 - requestTypeで読み取対象文書の種類を定義
 - 文書種類（預金通帳、請求書等）のみの選択
定義内容は固定

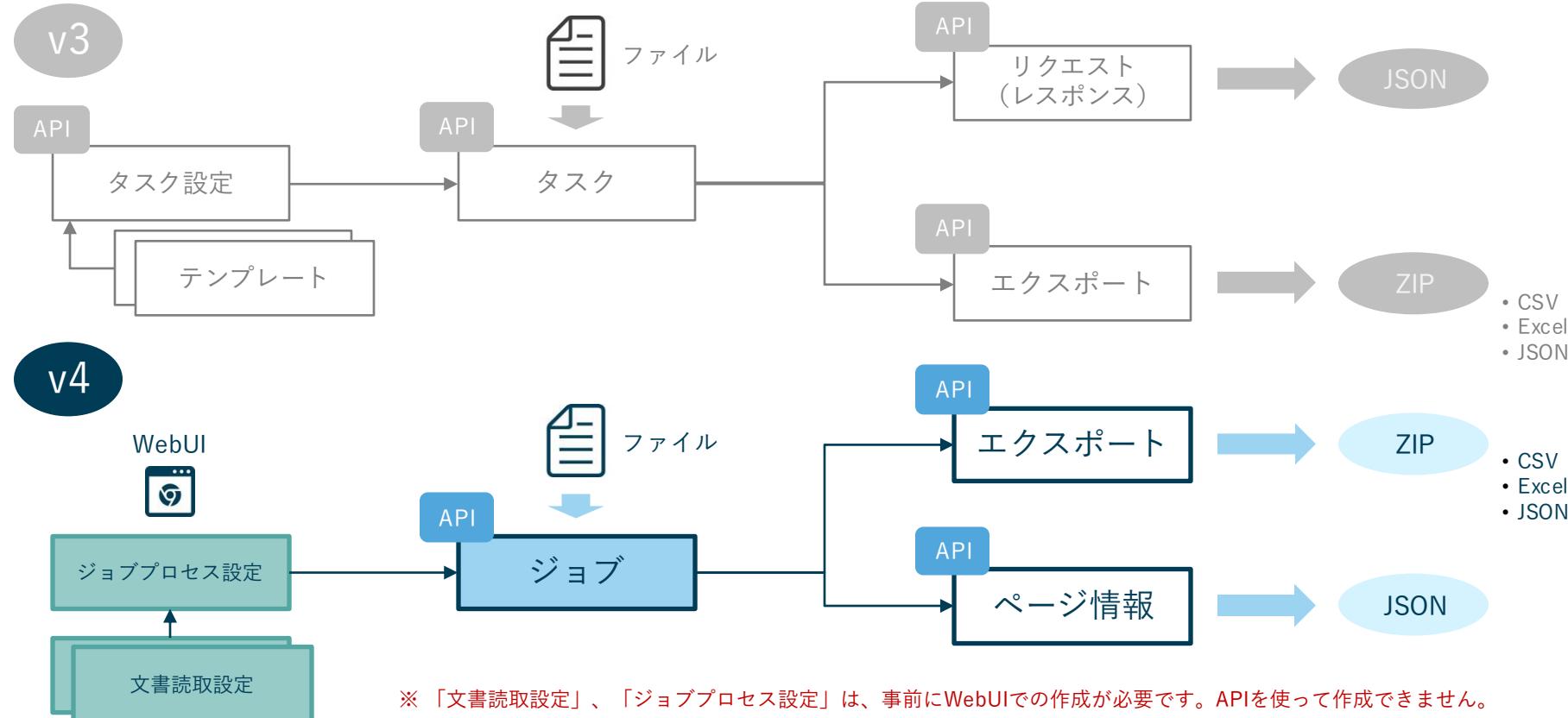
v4



- ジョブ

- ジョブ
 - 「文書読み取設定」、「ジョブプロセス設定」により、読み取箇所、出力内容のカスタマイズ性を向上
 - 出力項目の表示/非表示
 - 項目加工
 - 辞書

読み取処理概念の変更 (V3→V4) 2/2



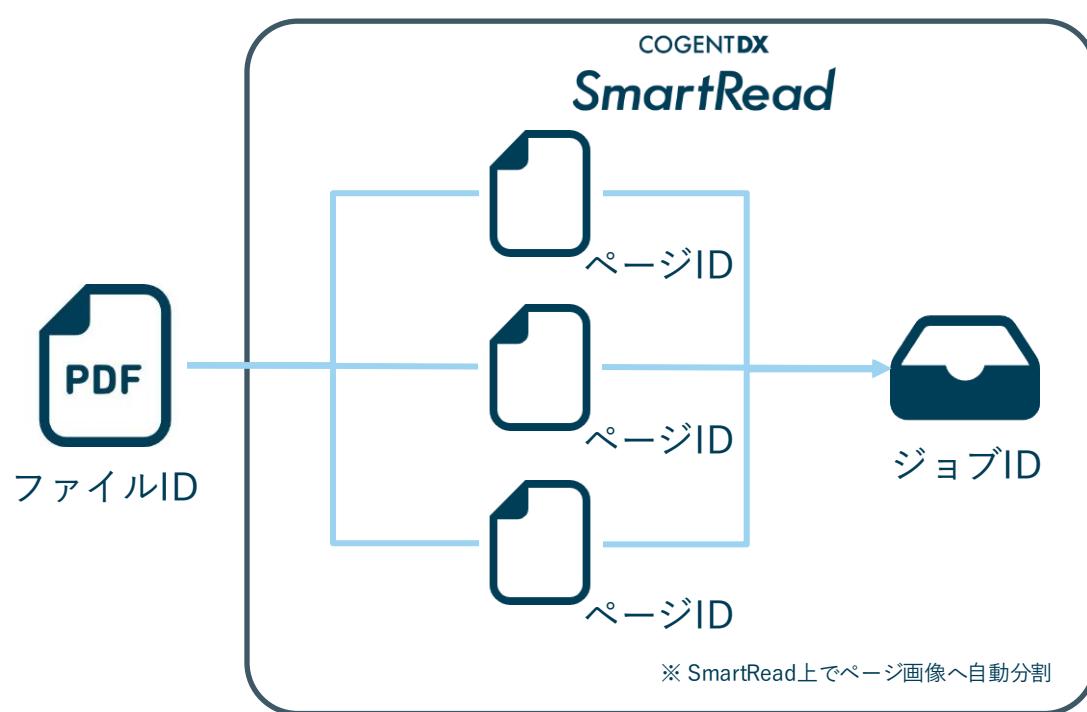
その他の変更 (V3→V4)



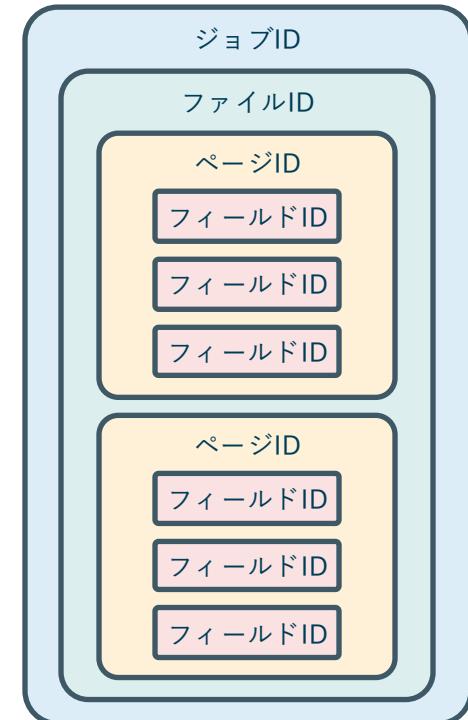
内容	V3	V4
読み取り対象文書 「給与支払報告書」	サポート終了	読み取り対象外
読み取り処理開始のタイミング	ファイル送信後、自動的に読み取り処理開始	明示的に「ジョブのスタート」が必要 (処理開始タイミングを明確化することで、 内部処理調査、今後の機能拡張を容易にする ことが狙い)
エクスポートのステータス	ステータス確認用エンドポイント有り	ステータス確認用エンドポイントは廃止 (ダウンロード用API実行時に、処理が 完了していればダウンロード、未完了な場合にはステータスが返ります。)
エクスポートによる JSONの出力内容	出力項目と読み取り結果に加え、 座標情報、確信度も出力	出力項目と読み取り結果のみ

※ 「座標情報」、「確信度」は、ページ情報取得用APIで取得可能です。

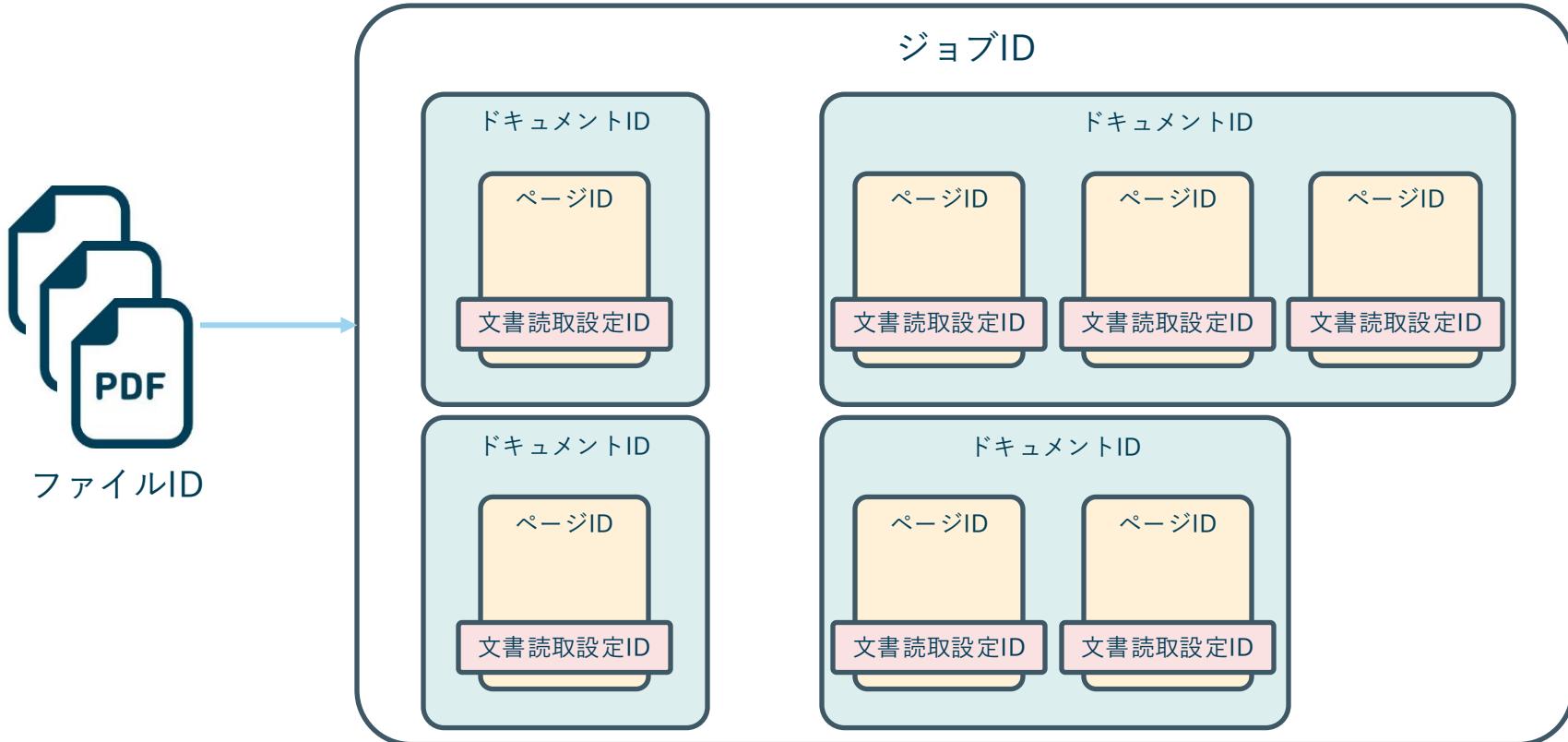
読み取ファイルに関するID 1/2



読み取対象ファイル投入の流れ



文書仕分けに関するID 2/2



※ SmartReadの「ドキュメント（ドキュメントID）」とは、処理に必要な情報が一式揃った塊」のことを指します。

目的別処理手順

1. 読取結果の取得（エクスポート）
2. ファイル画像データの取得
3. 項目画像データ, 座標情報, 確信度 の取得

再掲) 結果の取得方法と種類

出力方法 (エンドポイント)	出力 形式	読み取 り項目/結果	座標情報	確信度	読み取り 項目 画像取得URL	読み取り ページ 画像取得URL
/v4/export/:exportId ※ WebUI同様、ZIPファイルでの取得になります。	CSV Excel JSON	○				
/v4/verification-data/:jobId/results/:pageId	JSON	○	○	○	○	
/v4/verification-data/images/:fileId	JSON					○

※ 画像は直接取得できません。取得用URLにアクセスすることで取得できます。なお、URLの有効期限は1時間です。

共通の事前ステップ (= 読取処理の完了)

#	処理内容	WebUI	API (エンドポイント)
1	• 文書読み取り設定の作成 • ジョブプロセス設定の作成	○	※ APIを使って作成できません。
2	ジョブの作成	○	/v4/job
3	ファイルのアップロード	○	/v4/job/:jobId/file
4	ジョブの開始	○	/v4/job/:jobId/start
5	ジョブの確認 (ステータス)	○	/v4/job/:jobId
6	読み取り結果確認/修正作業	○	

ジョブの確認 (/v4/job/:jobId)

```
{
  "orgId": "alaad4de0-d624-4f73-9343-5aafb13d5935",
  "userId": "32981ee2-fffc-45a5-b208-092a7789a86a",
  "groupId": "15997db9-c2c9-4c96-a967-c5a1fela841a",
  "created": "2025-04-21T07:30:12.418Z",
  "modified": "2025-04-21T07:32:00.633Z",
  "deleted": null,
  "name": "サンプル仕分け",
  "description": null,
  "processId": "a8a3f914-98c8-4b82-8800-65ed06446f35",
  "usageGroupId": "15997db9-c2c9-4c96-a967-c5a1fela841a",
  "state": "COMPLETED",
  "allowExternalVerification": false,
  "externalVerificationComplete": false,
  "groupIds": [
    "15997db9-c2c9-4c96-a967-c5a1fela841a",
    "3e662bbe-2e98-48d1-bf25-88dba7729689"
  ],
  "pageStateSummary": {
    "CLASSIFICATION_RUNNING": 0,
    "CLASSIFICATION_VERIFIED": 0,
    "CLASSIFIED_AUTOMATICALLY": 0,
    "COMPLETED": 4,
    "FAILED": 0,
    "IGNORED": 0,
    "RUNNING": 0,
    "VERIFIED": 0
  },
  "numOfPages": 4,
  "fileStateSummary": {
    "CLASSIFICATION_RUNNING": 0,
    "CLASSIFICATION_VERIFIED": 0,
    "CLASSIFIED_AUTOMATICALLY": 0,
    "COMPLETED": 1,
    "FAILED": 0,
    "IGNORED": 0,
    "RUNNING": 0,
    "VERIFIED": 0,
    "CREATED": 0
  },
  "numOfFiles": 1,
  "verificationLockedBy": null,
  "files": [
    {
      "documentReaderId": "9f054999-84bb-4392-b93c-24ac971e8d9f",
      "documentReaderName": "サンプル: 請求書(明細)",
      "order": 1,
      "isSplit": false,
      "isAutoDetection": true,
      "original fileId": "11387f2f-89fd-4aa5-8f6f-234c9e6157",
      "fileId": "b20f60af-4b08-4214-be7f-2149af0bb4b7",
      "name": "sample_0008_01.pdf",
      "numOfPages": 4,
      "size": 65955,
      "type": "invoice",
      "state": "COMPLETED",
      "pages": [
        {
          "pageId": "0e869689-52ab-5eb9-be1b-fd0e3e7d5fd3",
          "index": 0,
          "size": 16488.75,
          "state": "COMPLETED",
          "verificationStatus": "UNVERIFIED"
        },
        {
          "pageId": "dae1da9c-6e08-546a-82dd-4f328f6a6256",
          "index": 1,
          "size": 16488.75,
          "state": "COMPLETED",
          "verificationStatus": "UNVERIFIED"
        }
      ]
    }
  ]
}
```

ジョブのステータス

```
"files": [
  {
    "documentReaderId": "9f054999-84bb-4392-b93c-24ac971e8d9f",
    "documentReaderName": "サンプル: 請求書(明細)",
    "order": 1,
    "isSplit": false,
    "isAutoDetection": true,
    "original fileId": "11387f2f-89fd-4aa5-8f6f-234c9e6157",
    "fileId": "b20f60af-4b08-4214-be7f-2149af0bb4b7",
    "name": "sample_0008_01.pdf",
    "numOfPages": 4,
    "size": 65955,
    "type": "invoice",
    "state": "COMPLETED",
    "pages": [
      {
        "pageId": "0e869689-52ab-5eb9-be1b-fd0e3e7d5fd3",
        "index": 0,
        "size": 16488.75,
        "state": "COMPLETED",
        "verificationStatus": "UNVERIFIED"
      },
      {
        "pageId": "dae1da9c-6e08-546a-82dd-4f328f6a6256",
        "index": 1,
        "size": 16488.75,
        "state": "COMPLETED",
        "verificationStatus": "UNVERIFIED"
      }
    ]
  }
]
```

ファイルID

ページID

目的別処理手順

1. 読取結果の取得（エクスポート）
2. ファイル画像データの取得
3. 項目画像データ, 座標情報, 確信度 の取得

取得結果 (JSON出力の場合)

```
[  
  {  
    "年": "3",  
    "月": "2",  
    "日": "15",  
    "摘要": "振込1",  
    "摘要2": "カ)エービ -シーショウ",  
    "お支払金額": "",  
    "お預り金額": "120,000",  
    "差引残高": "5,003,000",  
    "ファイル名": "通帳.pdf",  
    "ページ番号": "1"  
  },  
  {  
    "年": "3",  
    "月": "2",  
    "日": "15",  
    "摘要": "振込2"  
  }]
```

項目と読み取り結果というシンプルな出力形式

エクスポートの実行



#	処理内容	API (エンドポイント)
1	エクスポートの開始	/v4/export/:jobId
2	エクスポートのダウンロード	/v4/export/:exportId

目的別処理手順

1. 読取結果の取得（エクスポート）
2. ファイル画像データの取得
3. 項目画像データ, 座標情報, 確信度 の取得

取得結果

```
[  
  {  
    "pageID": "0e869689-52ab-5eb9-be1b-fd0e3e7d5fd3",  
    "imageUrl": "https://smartread-prod-requests.s3-ap-northeast-1.amazonaws.com/requests/b20f60a...  
    "pageIndex": 0  
  },  
  {  
    "pageID": "2ed96db3-ba0f-52c2-a40f-f9b45b7e49b5",  
    "imageUrl": "https://smartread-prod-requests.s3-ap-northeast-1.amazonaws.com/requests/b20f60a...  
    "pageIndex": 2  
  },  
  {  
    "pageID": "7ce03e8d-237c-5d82-98c2-e9c6f5ade7a4",  
    "imageUrl": "https://smartread-prod-requests.s3-ap-northeast-1.amazonaws.com/requests/b20f60a...  
    "pageIndex": 3  
  },  
  {  
    "pageID": "dae1da9c-6e08-546a-82dd-4f328f6a6256",  
    "imageUrl": "https://smartread-prod-requests.s3-ap-northeast-1.amazonaws.com/requests/b20f60a...  
    "pageIndex": 1  
  }  
]
```

ページ画像データ取得用URL

ファイルに含まれるページの画像データ取得用URLを確認

ファイル画像URLの取得

#	処理内容	API (エンドポイント)
1	画像取得用URLの取得	/v4/verification-data/images/:fileId

※ 「fileId」は、ジョブの詳細情報取得APIで取得可能です。

目的別処理手順

1. 読取結果の取得（エクスポート）
2. ファイル画像データの取得
3. 項目画像データ, 座標情報, 確信度 の取得

取得結果

```
{
  "fields": [
    {
      "fieldId": "80b2cea5-3b7c-49c1-a241-c0e35f749af",
      "apiVersion": "v4",
      "boundingRegion": [
        {
          "x": 2086,
          "y": 1298
        },
        {
          "x": 2086,
          "y": 1321
        },
        {
          "x": 2171,
          "y": 1321
        },
        {
          "x": 2171,
          "y": 1298
        },
        {
          "x": 2086,
          "y": 1298
        }
      ],
      "documentReaderFieldId": "b00cabd6-02ac-4c0c-861e-d852634d5b7a",
      "labelConfidence": 0.512788206206534,
      "name": "備考_notes",
      "verificationStatus": "unverified",
      "modified": "2025-04-22T01:17:09.427Z",
      "confidence": 0.9995025485337339,
      "value": "2200",
      "pageId": "147a79a6-0258-5000-8d0a-ec37b4196ee2",
      "outputName": "備考",
      "isContentTypeValid": true,
      "formattedOutput": "2200",
      "imageUrl": "https://smartread-prod-requests.s3-ap-northeast-1.amazonaws.com/requests/ce223870-9f43-4f11-8082-e7ea1517139b"
    },
    {
      "fieldId": "ce223870-9f43-4f11-8082-e7ea1517139b",
      "apiVersion": "v4",
      "boundingRegion": [
        {
          "x": 907,
          "y": 286
        },
        {
          "x": 731,
          "y": 287
        }
      ]
    }
  ]
}
```

座標情報

確信度

項目画像データ取得用URL

ページに含まれる読み取り結果情報の取得

ページ情報の取得



#	処理内容	API (エンドポイント)
1	ページ情報の取得	/v4/verification-data/:jobId/results/:pageId

※ 「pageId」は、ジョブの詳細情報取得APIで取得可能です。

文書仕分け用API

文書仕分け用API（エンドポイント）とその使用目的



エンドポイント	Method	使用目的
/v4/job/:jobId/document	GET	ジョブに含まれるドキュメント、ドキュメントに含まれるページのステータス情報を取得する
/v4/job/:jobId/document/:documentId/page/:pageId/exclude	POST	指定したページを該当ドキュメントから除外する
/v4/job/:jobId/document/:documentId/page/:pageId/include	POST	指定したページを該当ドキュメントに含める
/v4/job/:jobId/complete-classification-result-verification	POST	仕分け処理を完了する（仕分け確認完了の指示）

※ SmartReadの「ドキュメント（ドキュメントID）」とは、処理に必要な情報が一式揃った塊」のことを指します。

文書仕分けを行うには

- ジョブにファイルをPOSTする際、「isAutoDetection」で指定します。

```
v4/job/{{jobId}}/file
```

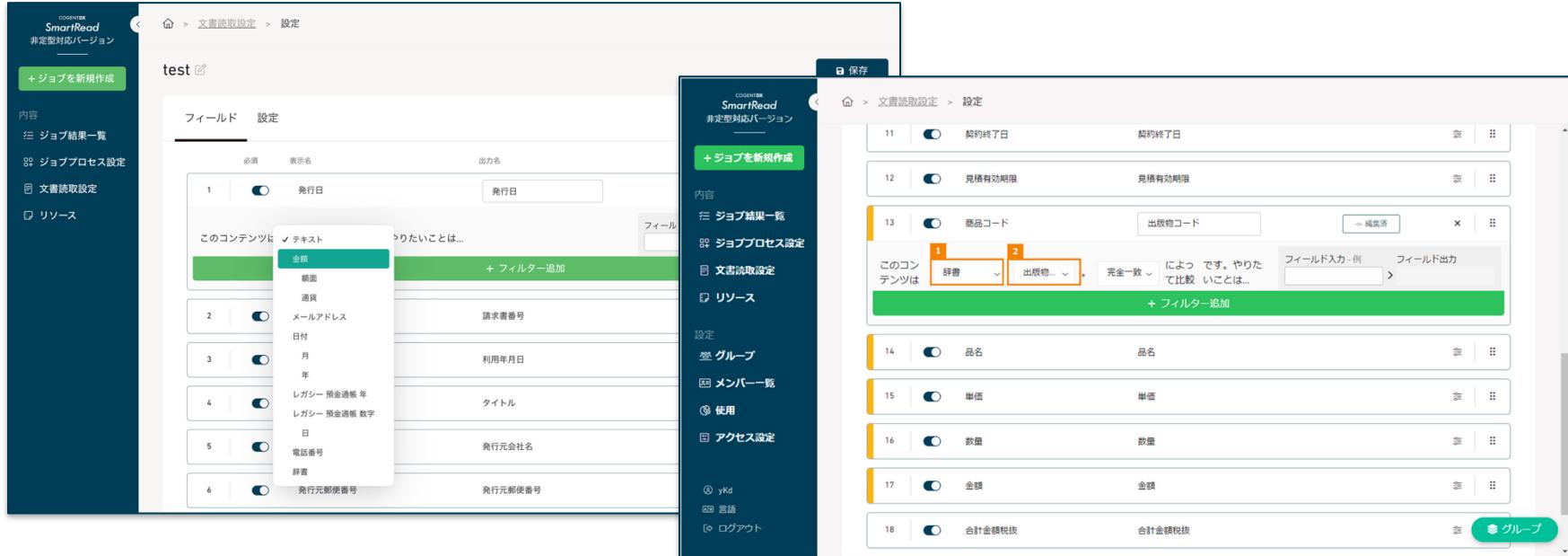
- documentReaderId
- image
- order
- isSplit
- originalFileId
- isAutoDetection**



「true」とすることで文書仕分けを実施
デフォルト値は「false」

参考

文書読み取り設定



The image displays the SmartRead document reading settings interface, specifically the 'Text' configuration screen. It shows a list of fields (1-18) with their corresponding labels and dropdown menus. A tooltip for '金額' (Amount) is visible, indicating it can be mapped to '請求書番号' (Invoice Number), '請求書' (Invoice), '出版物' (Publication), and '完全一致' (Exact Match). The interface is designed for flexible mapping based on document content.

これまでの1読み取り対象文書に固定設定ではなく、お客様の要件に合わせた柔軟な読み取り設定が可能

ジョブプロセス設定

SmartRead
非定型対応バージョン

+ ジョブを新規作成

内容

- ジョブ結果一覧
- ジョブプロセス設定
- 文書読み取り設定

リソース

設定

- グループ
- メンバー一覧
- 使用
- アクセス設定

- yKd
- 言語
- ログアウト

ジョブプロセス設定を新規作成

ジョブプロセス設定で使用される文書読み取り設定

+ 文書読み取り設定を選択

名前 ドキュメントの種類 ダウンロードするファイルの種類 説明

y_doc_見積書_03-1 見積書 .xls

ジョブプロセス設定を新規作成

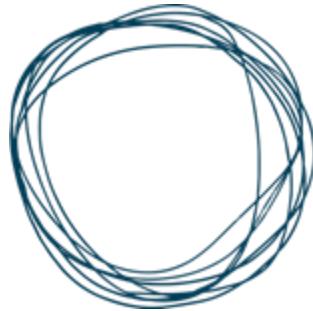
名前 y_proc_見積書_03-1

ダウンロードファイル形式

読取文書単位にファイルを出力
文書読み取り設定別にファイルを出力
文書タイプ(見積書、請求書など)単位にファイルを出力
文書タイプ(見積書、請求書など)ごとにシートを分けて出力

キャンセル

既存の「文書読み取り設定」を上書きすることなく、
必要に応じて読み取り処理毎の出力定義が可能



COGENT
LABS

免責事項

- ・ 本文書はご参考資料としてご提供するものです。本文書に関して弊社は質問へのご回答などサポートのご提供はいたしません。
- ・ 本文書は作成時において弊社が知り得た情報をもとに作成したものです。弊社は、本文書の正確性、網羅性、最新性などのいかなる事項についても保証いたしません。
- ・ 弊社は、本文書に関して、請求の原因の如何を問わず損害賠償責任その他の責任を負いません。