

# 実証実験報告書

## 防災FAX-OCR解析サービス実証開発

河川防災情報をデジタルの力で迅速に共有したい！

愛知県 建設局 河川課

×

協栄産業株式会社

# 目次

---



- 1) 実証開発の概要
- 2) 開発実証の結果
- 3) 課題と今後の展開
- 4) 参考資料

# 実証開発の概要

## ～防災FAX-OCR解析サービス～

- 背景・目的
- 全体概要



# 背景・目的



## 手作業による情報伝達作業

河川や海岸の水位情報をはじめとする「河川防災情報」を紙媒体に変換し、**手作業**で関係機関にFAXで伝達



## 情報伝達の遅延

河川防災情報が増えると伝達すべき機関が増えるため**情報伝達に遅延**が発生



## ヒューマンエラーの発生

担当者はマニュアルを参照しながら伝達先を選別する。河川防災情報が増えると**選別ミス**などの**ヒューマンエラー**が発生しやすくなる

関係機関から受信した河川防災情報について伝達すべき関係機関を自動判断

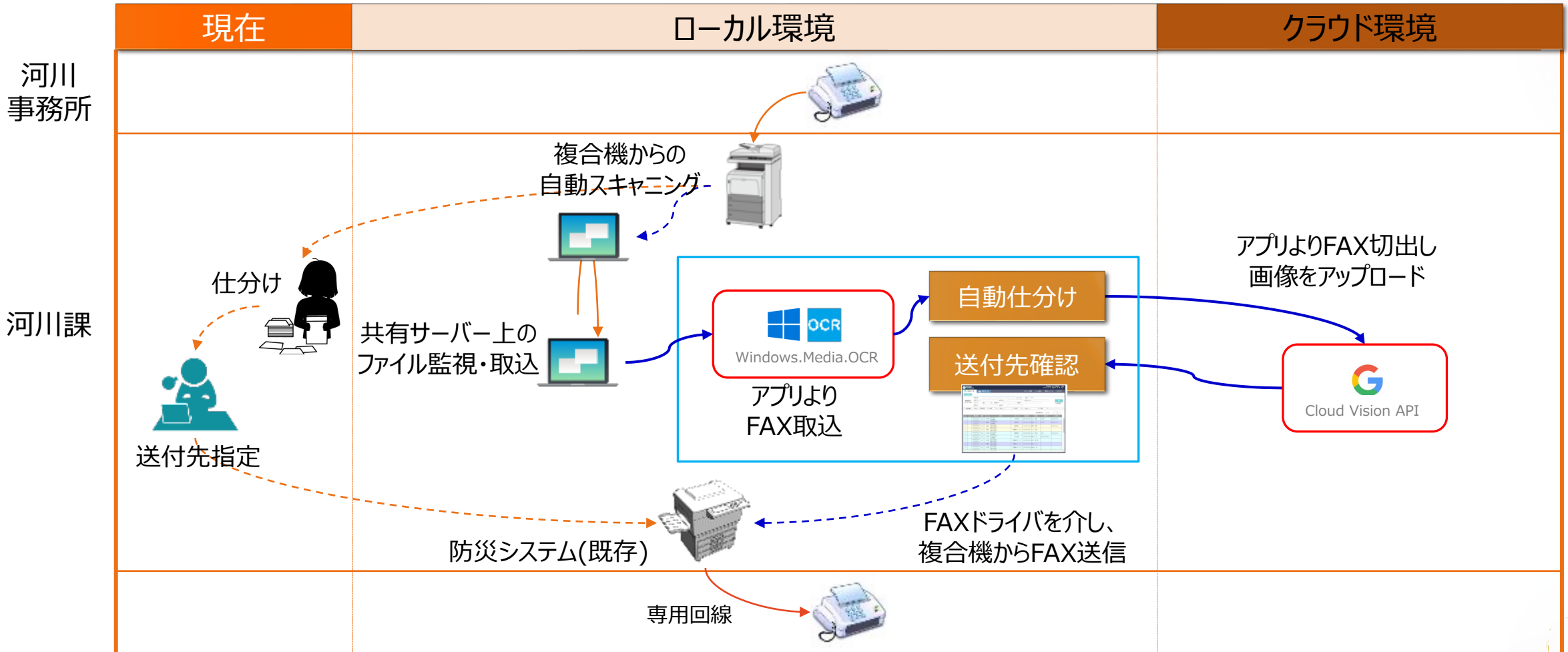


迅速かつ的確な河川防災情報伝達の実現

# 全体概要

## 運用イメージ

- - -> 手動・人による作業
- > FAX送信
- - -> 複合機機能にて実現
- > ローカル + クラウドサービス



アプリ(開発)

※Cloud Vision APIは、Google社のクラウドサービスです。

# 実証開発の結果

- ○ 目標指標 (KPI)
- ○ FAX-OCR解析サービスの特徴
- ○ 認識率・仕分けの検証に関して
- ○ 結果考察
- ○ 実証実験で得られた発見



# 指標目標 (KPI)

## 1. 業務の時間短縮

- ・FAX認識率：95%以上
- ・FAX仕分け／送信業務時間の削減：15%

## 2. FAX誤送信の防止

- ・FAX誤送信：0件 (100%)

## 3. 送信業務記録の保持

- ・送信業務操作ログ保存：100%

今回の実証開発で対象とした河川防災情報は、以下3つの予報・警報です

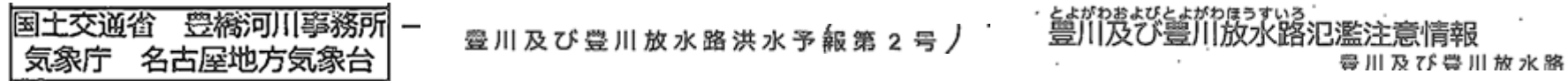
- a) 洪水予報      b) 水防警報      c) 津波水防警報



# FAX-OCR解析サービスの特徴

## FAX-OCRにて必要な主要機能の実装

a) 対象文字部分の画像切り出し



b) グレースケール処理 (文字画像をより滑らかにする処理)

c) 二値化処理 (黒をより黒く、白をより白くする処理)

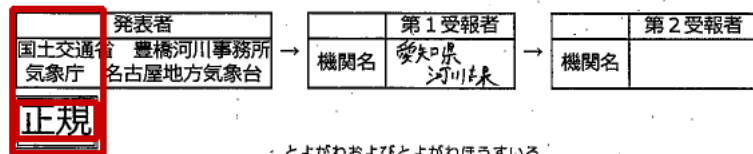
d) dpiスケールアップ (画像拡大: パラメータにて拡大率の指定が可能)

e) 誤認の多い文字への類似パターン登録 (圧と解析した場合、圧⇒庄として再解析する仕組み)

<主な類似解析候補>

国土交通省⇒国土交通省	: 土⇒士
木曾川中流⇒木曾川中流	: 曾⇒曾
庄内川⇒庄内川	: 圧⇒庄 等

f) 特定のマークを起点とした**相対位置から解析箇所**の文字解析



左記の例では、『正規』と記載されている箇所から相対位置を特定し、「発表者」のエリアを特定・文字解析を実施

g) 設定による仕分・解析

仕分方法や解析箇所の位置等、設定にて変更可能な構成としたため、**他の河川防災情報**や**他部署のFAX**解析も比較的容易に追加・変更が可能



# 認識率・仕分けの検証に関して

## 過去の河川防災情報（48パターン※1）を対象に4つのOCRにかけ、認識率の検証を実施

愛知県庄内川水系 新川氾濫注意情報

愛知県庄内川水系 新川洪水予報第〇号  
洪水注意報（発表）  
令和〇〇年〇〇月〇〇日〇〇時〇〇分  
愛知県尾張建設事務所・名古屋地方気象台 共同発表

（見出し）  
【警戒レベル2相当情報【洪水】】愛知県庄内川水系 新川では、氾濫注意水位に到達し、今後、水位はさらに上昇する見込み

（主文）  
【警戒レベル2相当】新川の水場川外水位観測所（清須市）では、「氾濫注意水位」に到達し、今後、水位はさらに上昇する見込みです。洪水に関する情報に注意して下さい。

（雨量）  
所により1時間に〇ミリの雨が降っています。  
今後もこの雨は降り続く見込みです。

流域	〇日〇時〇分～〇日〇時〇分までの流域平均雨量	〇日〇時〇分～〇日〇時〇分までの流域平均雨量の見込み
新川流域	〇〇〇ミリ	〇〇〇ミリ

（水位）  
愛知県庄内川水系 新川の水位観測所における水位は次のとおりと見込まれます。

観測所名	水位危険度 水位(m)	レベル1	レベル2	レベル3	レベル4
		水防団 待機	氾濫 注意	避難 判断	氾濫 危険
水場川外水位 水位観測所 (清須市)	00日00時00分の状況 xxx.x 00日01時00分の予測 xxx.x 00日02時00分の予測 xxx.x 00日03時00分の予測 xxx.x				

水位のグラフは各水位間を按分したものです。  
水位危険度レベル4については、氾濫危険水位と計画高水位を按分しており、氾濫危険水位＝計画高水位の場合は最大になります。なお、水位の予測値は前30分間の最大値を示しています。

（注意事項）

（参考資料）

（単位：水位（m））

観測所名	水場川外水位 水位観測所 清須市		
レベル4水位 氾濫危険水位*	5.20		
レベル3水位	4.40		

愛知県庄内川水系 新川氾濫注意情報

愛知県庄内川水系 新川洪水予報第〇号  
洪水注意報（発表）  
令和〇〇年〇〇月〇〇日〇〇時〇〇分  
愛知県尾張建設事務所・名古屋地方気象台 共同発表

（見出し）

【警戒レベル2相当情報【洪水】】愛知県庄内川水系 新川では、氾濫注意水位に到達し、今後、水位はさらに上昇する見込み

（主文）

【警戒レベル2相当】新川の水場川外水位観測所（清須市）では、「氾濫注意水位」に到達し、今後、水位はさらに上昇する見込みです。洪水に関する情報に注意して下さい。

（雨量）

所により1時間に〇ミリの雨が降っています。  
今後もこの雨は降り続く見込みです。

以下3つの認識率を導出  
仕分け・解析に必要な読取箇所

- ・文書タイトル
- ・表題
- ・送付元

※水防警報の解析には上記以外に「観測所」も読取箇所として用いています

Tesseract OCR	Windows.Media.OCR	Cloud Vision API	smart OCR
検証用アプリを開発し検証 <Python>	検証用アプリを開発し検証 <Python>	クラウドサービス利用し検証	試使用期間にて利用し検証 ※クラウドサービスを想定
<振分に必要な読取箇所> アプリによる領域指定 ※アプリによる座標指定	<振分に必要な読取箇所> 不可のため全文から導出	<振分に必要な読取箇所> 読取後、オブジェクトとして 自動分類されるため、対象 のオブジェクトを利用	<振分に必要な読取箇所> アプリによる領域指定 ※UIにて操作

※1河川防災情報（48パターン）は、『洪水予報』を対象として認識率の検証を実施



# 結果考察

## 1) 業務時間短縮 (FAX認識率)

OCR認識率向上は、業務時間短縮に大きく寄与するため、**95%以上の認識率を目指し**精度向上を実施

下記4つのOCR機能を用いて効果検証を実施

⇒**Google Vision API**が圧倒的に認識率が良い結果となった

ただし、将来的な運用を想定し、従量課金が必須のGoogle Vision APIを常に利用する事は避け、無償のMicrosoft.Media.OCRを用い、マッチングされない箇所のみGoogle Vision APIを利用する構成とし、認識率向上を実現

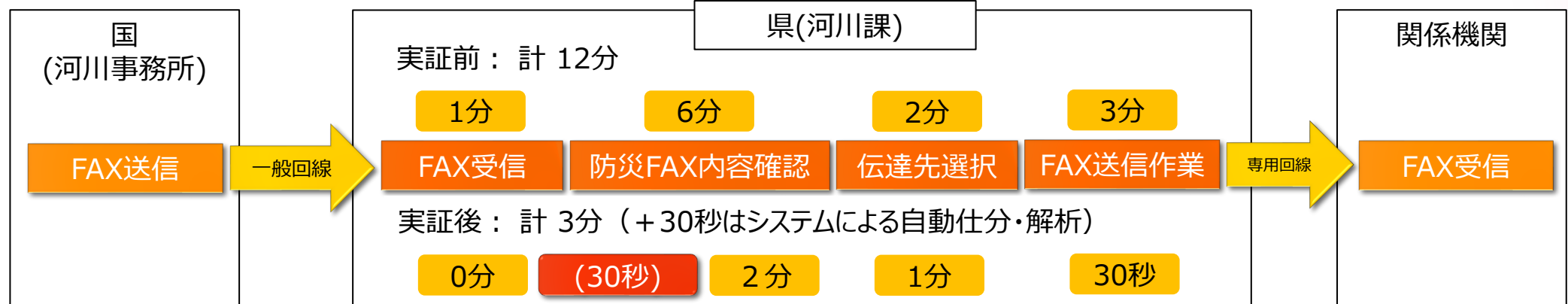
FAX認識率：**99.0%** (仕分け：**100%** / 解析率 **97.9%**)

	OCR	解析パターン			採用
		発令先	河川名	警報名	
ローカル：無償利用	Tesseract	52.1%	43.8%	12.5%	×
	Microsoft.Media.OCR	95.8%	60.4%	91.7%	○
クラウド：従量課金	Google Vision API	<b>100%</b>	<b>100%</b>	<b>93.8%</b>	○
ローカル or クラウド：有償	SmartOCR	100%	77% (認識パターン適用:98%)	98%	△

※ Microsoft.Media.OCR、Google Vision APIにおける認識率は、実証開発にて対象とした、『洪水予報』、『水防警報』、『津波水防警報』全ての認識率把握のため、継続して検証を行う事を予定しております

# 結果考察

## 2) 業務時間短縮 (仕分け・送信業務時間の削減)



解析件数：125件 (洪水予報:50件、水防警報:65件、津波水防警報:10件) / 平均解析時間：10.277秒/件

## 3) FAX誤送信の防止

認識率向上と共に、送信前に、解析結果・送信先が1つの画面で確認出来る画面を構築する事で、誤送信の防止に寄与 (今後の運用にて継続検証：**実証開発時点では誤送信 0件**)



# 結果考察

## 4) 送信業務記録の保持

データベースに記録する際、自動的に保存する仕組みを構築

※システム利用者、開発者共に操作ログの記録を意識せずに保持（保持率：**100%**）

## 5) その他、工夫／改善した点

検証前に想定してた認識率を大きく下回り、FAX-OCR特有の処理を様々なパターンを用いて検証

- ・対象文字部分の画像切り出し：FAX画像全体で解析するよりも**精度向上に寄与**
- ・グレースケール処理：**精度向上には大きく寄与せず**
- ・二値化処理：

**精度向上には多少寄与**したものの、FAX特有の情報落ちした箇所を**濁点等に判断する場合もあり**

- ・dpiスケールアップ（画像拡大）：  
**4～8倍**にスケールアップした場合、**大きく改善**したものの、**8倍以上とした場合、精度低下**
- ・誤認の多い文字への類似解析候補登録：実証開発時 **21パターン追加**

# 実証実験で得られた発見

## 1. 業務の効率化

これまで手作業で行っていた業務を自動化することで、効率性と情報伝達のスピードの向上に可能性を感じた。

## 2. 業務の可視化

実証実験を経て、業務の可視化や数値化ができた。また職員のスキルを正確に把握できるようになった。

## 3. 職員スキルのバラツキ解消

河川防災情報の伝達は、職員のスキルで伝達速度にバラツキがあった。今回の実証実験により、経験値に関係なく同じ処理速度で情報伝達が行えるようになった。



## 課題と今後の展開

- 今後の主な課題
- 今後のサービスの方向性

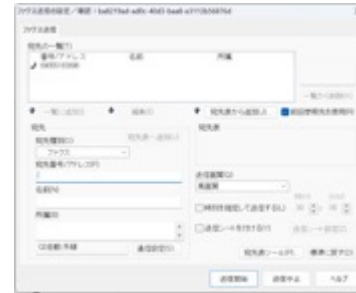
# 今後の主な課題

## 1) 複合機連携

複合機への送信に今回は複合機専用のFAXドライバーを用いたが、プレビュー画面※1、FAX架電先選択ダイアログ※2と、ドライバーが標準で保持している機能が立ち上がるため、操作として冗長なフローとなっている  
また、FAXドライバーへ受け渡した後の制御をアプリケーションから制御出来ない  
※複合機まできちんとデータ送信されたかが判定出来ない



※1 プレビュー画面



※2 FAX架電先選択ダイアログ

FAX送信ボタン押下後、FAXドライバーへ制御を渡すため、その先の操作情報はアプリケーションでは判断できない

サードパーティー製の複合機連携APIを用いる事で回避可能

## 2) 送信結果連携

FAX送信した結果は、愛知県既設の防災システムからFAXで返信されるが、正しく返信されたかの判断は人的に確認する必要があり、今後、FAX解析サービスで追加実現したい

愛知県既設の防災システムが関係機関へ再送する先をマスタとして登録し、返信FAXを解析する事で実装は可能

# 今後のサービスの方向性

<他自治体へ展開について>

愛知県庁

建設局 河川課

- ・実証開発における残課題の解決
- ・建設事務所における河川防災FAX業務のデジタル化
- ・多くのFAX業務でお困りの他部署のご紹介

他自治体向け対応

- ・FAX業務の効率化、自治体DX
- ・当社FAXソリューション（OCRエンジン）への導入
- ・自治体向けFAXソリューションとしてサービス化

愛知県庁におけるFAX業務課題に対し検討

様々なFAX業務へ対応可能なサービスへ

展示会、セミナー等、当社サービス製品としてPR

業務負荷を課題としている自治体へご提案

1) 更なる**自動化の推進**（複合機とのシームレス連携、OCR認識率向上、UI改善 等）

2) FAXを手放せない**自治体業務・部署向けサービスの検討**

3) 当社FAXソリューションの**OCRエンジン**（既存アプリ⇒Google Vision API）への**切り替え検討**

4) OCR技術を用いた**新たなサービスの検討**



# 参考資料

# OCRソフトウェア

## OCRソフトウェア対比表

ソフトウェア	開発元	主な特徴	実行環境	有償/無償	価格
Tesseract OCR	Google Inc. ※2005年まではヒューレット・パカード社により開発	ニューラルネットワークによるOCRエンジンを採用し、対応言語は116言語、認識率向上の各種機能も有している API利用が前提のため開発必須だが、個社対応が容易	ローカルPC サーバー	無償 ※オープンソースソフトウェア	—
Windows.Media.OCR	Microsoft Corporation	Windows10以降に無償で標準で導入されているRuntime API 認識率向上等の機能は有していないが、パソコンやエッジコンピューターとの親和性は非常に高い API利用が前提のため開発が必須	ローカルPC	無償 ※標準バンドル	—
Cloud Vision API	Google Inc.	機械学習を使用して、業界トップクラスの予測精度で画像を解析、また、カスタムレベルで画像を分類する機械学習モデルも実装 オブジェクトや顔を検出し、手書き文字の読取、有用な画像メタデータの抽出が可能	クラウドサービス	従量課金	全ての機能 ～1,000枚/月：無償 ドキュメントテキスト検出 1,001～5,000,000枚/月 1,000枚あたり \$1.50
smart OCR	株式会社インフォディオ	画像データからテキストデータを抽出する業務効率化支援ソリューションで、AIによるディープラーニングを活用し、従来のOCRでは難しかった、手書き文字や非定型帳票などにも対応	ローカルPC サーバー クラウドサービス	有償	オンプレミス：初期 100万円 ライセンス 10万円～/月 クラウド：初期 10万円 ライセンス 3万円～/月

# FAX-OCR解析結果

OCRパターン	発令先	河川名	警報名	採用
対象文字	国土交通省	木曽川中流 木曽川下流 長良川下流 庄内川 矢田川 矢作川 豊川及び豊川放水路	洪水予報	-
Tesseract	52.1%	43.8%	12.5%	×
Microsoft.Media.OCR	95.8%	60.4%	91.7%	○
Google Vision API	100%	100%	93.8%	○
SmartOCR	100%	77% (認識パターン適用:98%)	98%	△