

～ディープラーニングによる画像認識～  
**定型文書精確識別ソリューション**  
応用例：確定申告用紙識別

May, 2018



# 日本語文字OCRソリューションの現状

## ■ 市販ソフト

- 殆どWindows用の単体ソフトウェア。手動での操作が必要なため、業務システムに組み込むのは困難
- すべての文字を一括して認識。必要な部分だけ認識するのは困難
- 認識結果は人間が目視でチェックする必要があり、自動チェックは困難
- 認識精度に不満があっても、改善の余地はあまりない

## ■ Google Cloud Vision API

- APIを通して利用するものなので、業務システムに組み込み可能
- すべての文字を一括して認識。必要な部分だけ認識するのは困難
- 認識結果XMLを元にチェックツールを実装する余地はあるが、汎用的な機能であるが故、不要なテキストと必要なテキストが1つのテキストとして認識される場合があり、精度の高いチェックは困難
- 認識精度に不満があっても、改善の余地はない
- 画像をGoogleにアップロードする必要があるため、個人情報や秘密情報の外部サーバに流出恐れがあり

# 業務用定型文書認識の課題

- OCR対象画像の取り込み、OCR処理、処理結果のチェック、処理済み画像や読み込んだテキストの整理などの一連の処理を他システムと連携しながら自動化までをワンストップでやれば大きな業務効率改善になるが、それに見合ったソリューションは見たことがない
- 人間が行っている紙の情報からシステムへの転記作業を自動化したければOCR対象画像（文書）から必要な部分だけのテキスト化は必須
- 運用開始後もOCR対象画像のパターンは増えるため、継続的な機能改善や認識精度を向上させられる仕組みが必要

# 定型文書精確識別ソリューションの特徴

- ディープラーニングを活用して認識率を飛躍的に向上
- 認識できる内容：  
約 6800字（英字、数字、ひらがな、カタカナ、記号、第1水準漢字、第2水準漢字）
- フォント問わずに、どのフォントでも対応可能
- ご希望による特殊フォント・記号など随時追加対応は可能
- 内製サービス提供の形で、外部サービスを依存せず、現行システムとの連携は自由自在
- 定型文のフォーマットに合わせる、結果のチェックツールを提供、チェック要員の負担を軽減
- 定型文の内容に合わせる、合計数字の照合などロジックのチェック機能を付加可能
- 誤認識あっても、結果チェックツールに素材を貯蓄して、ディープラーニング訓練に投入して、認識ライブラリを更新により、認識ができるようになる

# ソリューション別機能比較表

	システム連携	識別率	誤認識対策 (認識率向上)	精確認識 (関心部分のみ 改行など対応)	結果チェックツール	特殊フォント 記号拡張対応	結果自動照合
市販ソフト	×	○	×	×	×	×	×
Google Cloud Vision API	△	○	×	×	△	×	×
定型文精確識別 ソリューション	○	○	○	○	○	○	○

- 対応可能
- △ 対応困難
- × 対応できない

# 定型文書精確識別ソリューション

## 応用例

給与所得者の保険料控除申告書

# 読み込み対象定型文書

## 少し斜めに読み込まれた「給与所得者の保険料控除申告書」

### 平成30年分 給与所得者の保険料控除申告書

所轄税務署長 千代田区 税務署長	給与の支払者の名称(氏名) 任天堂 株式会社	あなたの住所 東京都中央区日本橋堀留町2丁目6-9
	給与の支払者の法人番号 0123456789012	あなたの氏名 田中 太郎
	給与の支払者の所在地(住所) 東京都中央区小伝馬町4-2 VORT日本橋本町9階	あなたの住所 東京都中央区日本橋堀留町2丁目6-9



保険会社等の名称	保険等の種類	保険期間又は支払期間	保険等の契約者の氏名	保険金等の受取人の氏名	新・旧の区分	あなたが本年中に支払った保険料等の金額(控除対象金額)	給与の支払者の確認印
東京海上日動	生命教育	2年	吉行 花子	安井 麻耶 本人	新	1,234,567,890	(a)
三井住友海上	健康交通	5年	松田 一郎	高橋 太陽 妻	旧	12,345	(a)
あいおいニッセイ同和損保	火災家財	2016/02	豊田 和子	東尾 美奈代 夫	新	238,781,324	(a)
損保ジャパン日本興亜株式会社	生活のための	2008/12	菅田 海老蔵	堀ちえみ 次女	旧	3,678,612,936	(a)
Aのうち新保険料等の金額の合計額		A		計(①+②) ③		12,361円	
Bのうち旧保険料等の金額の合計額		B		計(③+④) ⑤		12,378円	
アクサ生命保険株式会社	医療保険	期間	山崎 将暉	戸田 由里子 長男	新	7,865,234,572	(a)
FWD富士生命	引当基礎保証	年払	長澤 さとみ	上野 はるか	新	23,469	(a)
Cの金額の合計額		C		計(⑤+⑥) ⑦		36,934円	
SBI損保の火災保険	自分でもらえる	特別適用	満島 千鶴	新	9,076,982,345	(a)	
セゾン自動車火災保険	基本の補償	権利完全	貫地谷 恭子	旧	89,263,489	(a)	
あいおいニッセイ同和損保	損害防止費用	行使	吹石 祐実	新	74,761	(a)	
Dのうち新保険料等の金額の合計額		D		計(⑦+⑧) ⑨		23,689円	
Eのうち旧保険料等の金額の合計額		E		計(⑨+⑩) ⑪		23,624円	
計算式Ⅰ(新保険料等) ※		計算式Ⅱ(旧保険料等) ※		生命保険料控除額計(⑨+⑩+⑪)		118,378円	

保険会社等の名称	保険等の種類	保険期間	契約者の氏名	あなたが本年中に支払った保険料等の金額(控除対象金額)	給与の支払者の確認印	
分譲マンション基準料	家財安心プラン	3年	石原 エリカ	地震	2,345,682,734	
津波を原因とする火災補償する	実際修理費	5年間	広末 宏美 長男 寺西 充希 土屋 美玲 長女	旧長期	423,458	
Aのうち地震保険料の金額の合計額		A		計(①+②) ③		23,456,745
Bのうち旧長期損害保険料の金額の合計額		B		計(③+④) ⑤		234,575
Cの金額(⑤の金額が10,000円を超える場合は(⑤×1/2-5,000円) ※)		C		計(⑤+⑥) ⑦		23,412

社会保険の種類	保険料支払先の名称	保険料を負担することになっている人	あなたが本年中に支払った保険料の金額
船員共済組合	退職者医療制度	米倉 美穂 祖母	8,268,346,238
国民厚生	厚生年金保険の適用を受ける会社に勤務する	松下 美帆 叔父	78,962,346
合計(控除額)			2,873,648,263

種類	あなたが本年中に支払った掛金の金額
独立行政法人中小企業基盤整備機構の共済契約の掛金	7,342,578
確定拠出年金法に規定する企業型年金加入者掛金	78,623,641
確定拠出年金法に規定する個人型年金加入者掛金	8,688,971,234
心身障害者扶養共済制度に関する契約の掛金	65,761,234
合計(控除額)	7,856,783,415

◎ この申告書の記載に当たっては、裏面の説明をお読みください。

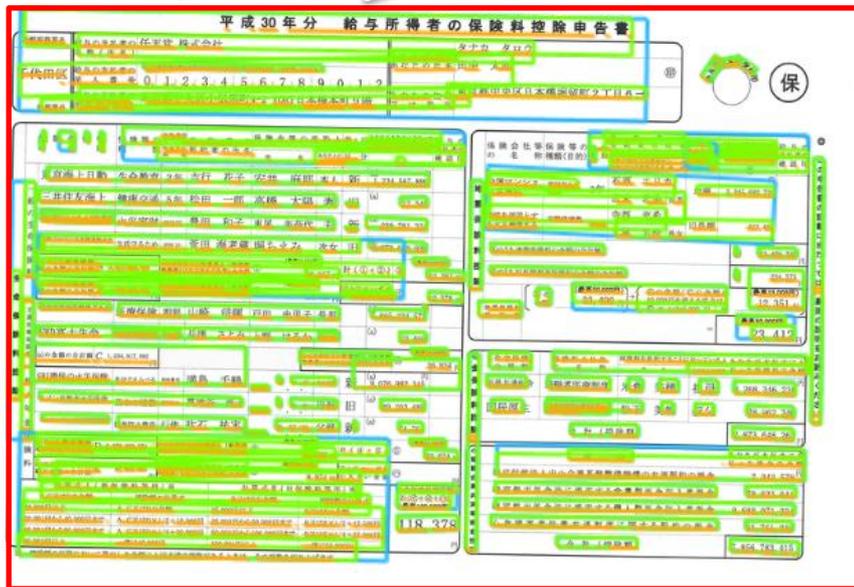
※ 控除額の計算において算出した金額に1円未満の端数があるときは、その端数を切り上げます。



# Google Cloud Vision APIの識別結果



すべての文字を識別



実際近い例\_斜め\_600.png

フォーマットと記入したテキストの切り分けができない

+Block 3

平成30年分給与所得者の保険料控除申告書 所轄税務署長「給与の支払者の任  
天堂株式会社

タナカ タロウ 名称 (氏名)

「タナカ タロウ 名称 (氏名)」として認識してしまい、ここから「タナカ タロウ」だけを取り出すことは困難

支払者の ※この申告書の提出を受けた給与の「法人番号 0 | 1 | 2 | 3 | 4 | 5 | 1  
67 | 8 | 9 0 : 1 | 12

に「東京都中央区日本橋堀留町2丁目6  
- 9

# Google Cloud Vision APIの識別結果



+Page 1 +Block 1

老受水

+Block 2

6 2 P S

+Block 3

平

成30年分給与所得者の保険料控除申告書所轄税務署長「給与の支払者の任天堂株式会社

タナカ

タロウ名称(氏名)

(

フリガナ)上

あなたの氏名

田中太郎千代田区給与の支払者の※この申告書の提出を受けた給与の「法人番号0|1|2|3|4|5|167」8|90:112

に

「東京都中央区日本橋堀留町2丁目6-9

あなたの

住所沢給与の支払者の東京都中央区小伝馬町4-2VORT日本橋本町9階|

不備牛円

3階収は居所税務署長所在地(住所)

+Block 4

\*

040

+Block 5

※

この申告書の提出を受けた給与の支払者(個人を除きます。)が記載してください。

+Block 6

保

+Block 7

險

会

名

+Block 8

社

+Block 9

等

称

+Block 10

保険

の

あなたが本年中に支払った

契約者の氏名保険

又は

旧長期

保険料等のうち

、左欄の区分間|保険等の対象となった家あなた

に係る金額

(分配を受けた剰支払者の

余金等の控除後の金額

)屋等に居住又は家財をとの

確

\*Other names and brands may be claimed as the property of others.

Copyright © 2013 BITWARE Corporation. All rights reserved. BITWARE, the BITWARE logo, and BITWARE are trademarks of BITWARE Corporation in the Japan and other countries.





# 定型文精確認識ソリューション - チェックツール実装例①



東京都中央区小伝馬町4-2 VORT日本橋本町 9階

東 → 東 東 東 束 夷  
0.658 0.322 0.020 0.000 0.000

京 → 京 哀  
1.000 0.000

都 → 都 郡 部 邦 都  
0.960 0.040 0.000 0.000 0.000

中 → 中 申 甲 巾 市  
1.000 0.000 0.000 0.000 0.000

央 → 央 央 失 尖 夾  
0.998 0.002 0.000 0.000 0.000

区 → 区 反 匹 庄 医  
1.000 0.000 0.000 0.000 0.000

小 → 小 か 八 ぶ 亦  
1.000 0.000 0.000 0.000 0.000

認識候補TOP5を表示します

誤認識した場合でも、認識候補から1クリックで修正可能  
同時に誤認識素材として自動登録、次回ライブラリ更新により誤認識がなくなる



# 定型文精確認識ソリューション - チェックツール実装例②



種類	あなたが本年中に支払った掛金の金額
独立行政法人中小企業基盤整備機構の共済契約の掛金	7,342,578
確定拠出年金法に規定する企業型年金加入者掛金	78,623,641
確定拠出年金法に規定する個人型年金加入者掛金	8,688,971,234
心身障害者扶養共済制度に関する契約の掛金	65,761,234
合計 (控除額)	7,856,783,415

65,761,234
6,000
5,000
7,000
6,000
1,000
2,000
3,000
4,000
7,000
8,000
5,000
6,000
7,000
8,000
3,000
4,000
1,000
5,000

予め定義することで、明細項目と合計項目の数値を読み取り、明細項目の数値の合計が合計項目の数値と一致しているか自動的にチェックできます



# 定型文精確認識ソリューション - チェックツール実装例③

## 特別費用

平成 30 年 分 給 与 所 得 者 の 保 険

所轄税務署長 千代田区 税務署長	給与の支払者の名称(氏名) 任天堂 株式会社	給与の支払者の法人番号 0123456789012	給与の支払者の所在地(住所) 東京都中央区小伝馬町4-2 VORT日本橋本町9階	あなたの氏名 (フリガナ) 田中 フ	あなたの住所 又は居所 東京都 タナカ
------------------------	---------------------------	------------------------------	---	--------------------------	------------------------------

保険会社等の名称	保険等の種類	保険期間	契約者の氏名	保険金等の受取人	新・旧区分	あなたが本年中に支払った保険料等の金額(分配を受けた親身金等の控除後の金額)	給与の支払者の印
東京海上日動	生命教育	2年	吉行 花子	安井 麻耶 本人	新 (a)	1,234,567,890 円	
三井住友海上	健康交通	5年	松田 一郎	高橋 太陽 妻	旧 (a)	12,345 円	
あいおいニッセイ同和損保	火災家財	2018/02	豊田 和子	東尾 美奈代 夫	新 (a)	238,781,324 円	
損保ジャパン日本興亜株式会社	生活守るため	2028/12	菅田 海老蔵	菅田 海老蔵	新 (a)	612,936 円	
(a)のうち新保険料等の金額の合計額 A						2,871,023.4 円	(最高40,000円)
(a)のうち旧保険料等の金額の合計額 B						238,476.1 円	
A+C又はDの金額						3,109,499.5 円	
B又はEの金額						12,378 円	
計(④+⑤)						3,121,877.5 円	
介護医療保険料	FWD富士生命	引受基礎額和型 平成28年	大上野 はるか	妻の父親	新 (a)	23,469 円	
(a)の金額の合計額 C						1,234,617,892 円	
Cの金額を下の計算式Ⅰ(新保険料等)に当てはめて計算した金額						36,934 円	
個人年金保険料	SBI損保の火災保険	自分でえらべる	満島 千鶴	妻の母親	新 (a)	9,076,982,345 円	
(a)のうち新保険料等の金額の合計額 D						2,378,892,374 円	
(a)のうち旧保険料等の金額の合計額 E						87,972,345 円	
Dの金額を下の計算式Ⅰ(新保険料等)に当てはめて計算した金額						23,689 円	
Eの金額を下の計算式Ⅱ(旧保険料等)に当てはめて計算した金額						8,974 円	
計(⑥+⑦)						23,624 円	
A,C又はDの金額						20,000円以下	
B又はEの金額						25,000円以下	
控除額の計算式						25,000円以下	
A,C又はDの金額						20,001円から40,000円まで	
B又はEの金額						25,001円から50,000円まで	
控除額の計算式						25,001円から50,000円まで	
A,C又はDの金額						40,001円から80,000円まで	
B又はEの金額						50,001円から100,000円まで	
控除額の計算式						50,001円から100,000円まで	
A,C又はDの金額						80,001円以上	
B又はEの金額						100,001円以上	
控除額の計算式						一律に50,000円	
生命保険料控除額						118,378 円	

小さい文字でも精確認識できます

保  
の  
分  
地  
震  
保  
険  
料  
控  
除  
地  
区  
控  
社  
会  
保  
険  
料  
控  
除  
小  
保  
険  
企  
業  
共  
済  
等  
金  
庫  
控  
除

特	→	特	0.951	持待特時
別	→	別	1.000	則刑即川
費	→	費	0.998	贅賽贅寶
用	→	用	1.000	円印吊甲

\*Other names and brands may be claimed as the property of others.  
Copyright © 2013 BITWARE Corporation. All rights reserved. BITWARE, the BITWARE logo, and BITWARE are trademarks of BITWARE Corporation in the Japan and other countries.





# 定型文精確認識ソリューション – 前提条件



- スキャナーの読み取り解像度は600dpi以上

※600dpi以下の場合、認識精度下がる可能性があります

お問合せ

# ビットウェア株式会社

電話番号 : 03-6661-6920

FAX : 03-6661-6921

Web : <http://www.bit-ware.co.jp/>

メール : [info@bit-ware.co.jp](mailto:info@bit-ware.co.jp)

〒103-0001

東京都中央区日本橋小伝馬町4-2

VORT日本橋本町9階

# Thanks a lot

---

