

## クラウドソーシングにおける管理者作業の軽減に向けた試み

寺口 正義<sup>†</sup>      齋藤 新<sup>†</sup>      大野 正樹<sup>†</sup>

クラウドソーシングで複雑な作業を行う場合には管理者作業、つまり対象となるドメインにおいて高い知識を持つ人間による作業が不可欠である。本発表では、手書きシートの書き起こし支援システムを対象に、管理者作業を適切に分割し、クラウドソーシングおよび既存の自然言語処理技術を活用することで、ドメインに関する知識をあまり持ち合わせていない一般の作業者を利用した場合でも同様の目的を達成できるかを実験により考察する。

### Trial of Reducing Manager's Tasks in Crowdsourcing

MASAYOSHI TERAGUCHI<sup>†</sup>      SHIN SAITO<sup>†</sup>      MASAKI OHNO<sup>†</sup>

It is essential for experts who have high level of domain knowledge to do complicated tasks in crowdsourcing, called as manager's tasks. In this paper, we propose a new trial of reducing manager's tasks in crowdsourcing by dividing the tasks into appropriate small tasks and assigning them to general people who has less domain knowledge. We also discuss if even the general people can achieve the purpose of tasks properly and if their output have much higher quality than the result automatically generated by a machine with existing natural language processing technologies.

#### 1. はじめに

2011年3月11日に発生した東北地方太平洋沖地震とそれに伴う津波により、未曾有の東日本大震災が発生し、30万人を超える被災者を出した[1]。緊急の状況下において、国や地方自治体などは震災発生直後から、手書きの紙、画像、音声、映像、PDFといった様々なメディアタイプを通じて様々な種類の情報（被災者の安否、支援物資の配給、公共交通機関の復旧および行政サービスの状況、仮設住宅の申し込みなど）を迅速に公開してきた。

しかし、震災発生後しばらくの間、被災者の多くは避難所で不慣れな集団生活をしながら、非常に限られた手段を駆使して必要な情報を収集しなければならなかった。例えば、電力不足、電話回線の切断、無線基地局の故障等により電話や携帯メールが使えない状態が震災後しばらく続いた[2]ため、自分の足を使って情報を収集したり、TVやラジオといったマスメディアを介して一斉配信される情報を信じたりするしかなかった。

被災地での携帯基地局および電力の復旧に伴い、携帯電話によるインターネットアクセスが被災者の情報収集に役立つようになった。特にTwitter, Mixi, Facebookに代表されるソーシャル・ネットワーク

グ・サービス(SNS)が、携帯電話で安否情報や支援物資情報等を収集するための有効なツールとなった[2]。震災時にSNSで構築された社会的ネットワークが有効に機能したことも影響し、被災地内外の人々がネットワークを通じて協力しあい、特定の作業を行うクラウドソーシングが注目され始めた。

筆者らも震災復興支援の一環として、壊滅的な被害を受けた被災地に位置するとある町（以下、A町とする）の社会福祉協議会に対して、クラウドソーシングを活用して、手書きシートの書き起こしを支援するシステムを無償提供した。現地でのボランティア活動を通じて、また現場スタッフへのヒアリングに基づいて実際に必要とされるニーズを正しく捉えることで、現地と書き起こしボランティアの橋渡しには成功したが、開発期間が短く管理者視点での作りこみが満足にできなかったため、結果として手書きシートの書き起こし準備のための管理者の作業負担が増した。

本発表では、筆者らが無償提供したシステムの管理者作業のうち、シートに記載されている手書き情報の意味付けに注目し、その意味付けにクラウドソーシングおよび自然言語処理を活用することで管理者作業の軽減につながるかを、管理者作業量および書き起こし結果、書き起こされたデータの検索精度の点から考察する。

<sup>†</sup> 日本アイ・ビー・エム株式会社 東京基礎研究所  
IBM Research – Tokyo, IBM Japan Co., Ltd.

仮設住宅生活者実態調査票(\*\*\*ボランティアセンター) 訪問日: / /

|             |    |    |     |
|-------------|----|----|-----|
| 調査員氏名       | 氏名 | 住所 | 訪問者 |
| 氏名(仮設者)     | 氏名 | 住所 | 訪問者 |
| 元々の住所       |    |    |     |
| 家族構成        |    |    |     |
| 世帯主         |    |    |     |
| 世帯主の氏名      |    |    |     |
| 世帯主の住所      |    |    |     |
| 世帯主の電話番号    |    |    |     |
| 世帯主のメールアドレス |    |    |     |
| 世帯主の職業      |    |    |     |
| 世帯主の収入      |    |    |     |
| 世帯主の支出      |    |    |     |
| 世帯主の貯蓄      |    |    |     |
| 世帯主の資産      |    |    |     |
| 世帯主の負債      |    |    |     |
| 世帯主の健康      |    |    |     |
| 世帯主の生活      |    |    |     |
| 世帯主の社会      |    |    |     |
| 世帯主の文化      |    |    |     |
| 世帯主のその他     |    |    |     |

仮設住宅生活者実態調査票(\*\*\*ボランティアセンター) 訪問日: / /

|             |    |    |     |
|-------------|----|----|-----|
| 調査員氏名       | 氏名 | 住所 | 訪問者 |
| 氏名(仮設者)     | 氏名 | 住所 | 訪問者 |
| 元々の住所       |    |    |     |
| 家族構成        |    |    |     |
| 世帯主         |    |    |     |
| 世帯主の氏名      |    |    |     |
| 世帯主の住所      |    |    |     |
| 世帯主の電話番号    |    |    |     |
| 世帯主のメールアドレス |    |    |     |
| 世帯主の職業      |    |    |     |
| 世帯主の収入      |    |    |     |
| 世帯主の支出      |    |    |     |
| 世帯主の貯蓄      |    |    |     |
| 世帯主の資産      |    |    |     |
| 世帯主の負債      |    |    |     |
| 世帯主の健康      |    |    |     |
| 世帯主の生活      |    |    |     |
| 世帯主の社会      |    |    |     |
| 世帯主の文化      |    |    |     |
| 世帯主のその他     |    |    |     |

生活支援シート(世帯の基本情報)

世帯主: 氏名 住所 電話番号

世帯員: 氏名 住所 電話番号

世帯主の職業: 氏名 住所 電話番号

世帯主の収入: 氏名 住所 電話番号

世帯主の支出: 氏名 住所 電話番号

世帯主の貯蓄: 氏名 住所 電話番号

世帯主の資産: 氏名 住所 電話番号

世帯主の負債: 氏名 住所 電話番号

世帯主の健康: 氏名 住所 電話番号

世帯主の生活: 氏名 住所 電話番号

世帯主の社会: 氏名 住所 電話番号

世帯主の文化: 氏名 住所 電話番号

世帯主のその他: 氏名 住所 電話番号

図1 仮設住宅生活者向けの世帯調査シートの変遷。左が当初の巡回時のもので、真ん中を経て、右が現在のもの。

## 2. クラウドソーシングを活用した手書きシートの書き起こし支援システム

壊滅的な被害を受けた被災地に位置するとある町(以下、A町とする)の社会福祉協議会に対して、当社は震災直後から継続的な支援を続けている。著者らは、支援の一環として7月に行われた現地へのボランティアバスツアーに参画してボランティアとして活動する傍ら、現場での被災者のケアについての現状把握と問題点の洗い出し、情報通信技術を使った有効的な支援の検討を行った。その結果、著者らもクラウドソーシングを活用した書き起こし支援システムを開発し、ボランティアセンターを運営する社会福祉協議会に無償提供する運びとなった。

### 2.1 社会福祉協議会と災害ボランティアセンター

社会福祉法人 社会福祉協議会は民間の社会福祉活動を推進することを目的とした民間組織で、各地方自治体と連携して地域の福祉活動を担っている。平常時はこの社会福祉協議会が各地域のボランティアセンターを運営しているが、災害時には彼らが母体となって「災害ボランティアセンター」を設立・運営し、各地から集まるボランティアと共に、災害復旧・復興のための活動を行う。

### 2.2 被災者の心のケア

津波による壊滅的な被害を被った地域では、仮設住宅での暮らしを余儀なくされている被災者が非常に多い。特に震災および津波で家族を失った被災者が心に負った傷は非常に大きく、時間の経過に伴い、家族を失った寂しさや何も変わらない現状に対する無力感、

近隣に知人がいない孤立感などから自殺や孤独死してしまう人がある[3]。このような状況に陥らないよう、被災地の行政および社会福祉協議会は被災者宅を定期的に訪問して被災者の健康状態や精神状態等を調査し、必要に応じて継続的な心のケアを続ける必要がある。

### 2.3 被災者の心のケアにおける課題

A町では社会福祉協議会およびボランティアセンターが中心となり、被災者のケアを進めている。被災者宅を訪問して得られた調査結果は生活調査シートと呼ばれるシートに手書きで記入され、バインダーに綴じて管理されている。一部はExcelファイルに書き起こされて電子化されてはいたが、情報の検索や分析といった用途で有効活用するにはいくつかの問題があった。例えば、電子化されたデータは全ての被災者宅を網羅しているわけではないため、どの被災者宅を定期的に訪問すべきかを検討したり、過去の訪問履歴を確認したりするためには、Excelで検索しつつもバインダー片手に調べなければいけない状況であった。

#### 2.3.1 調査シートの変遷

震災当初は、社会福祉協議会が独自に作成した調査シートを活用して被災者宅を巡回していたが、最近では、全国社会福祉協議会が提供する全国共通フォーマットをベースにした調査シートを活用して、定期的な巡回が必要な世帯のみを巡回できるようになった。

図1に、一例として仮設住宅生活者向けの世帯用調査シートの変遷を示す。シートの変更に応じて調査項目が追加、修正され、詳細化されたことが分かる。シートに記載された手書き情報を電子化し、検索や分析といった用途で有効活用するには、それぞれの情報の

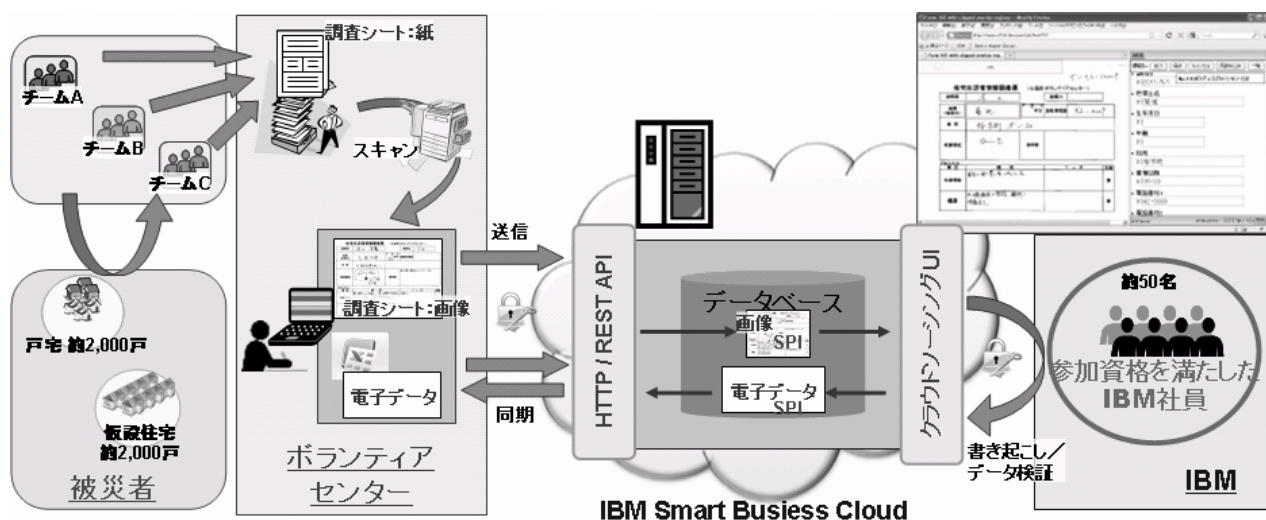


図2 書き起こし支援システムの全体概要図

意味を正しく認識できなければならず、そのような意味付けの作業には巡回訪問に関するドメイン知識を持った専門家の助けが必要となる。

### 2.3.2 膨大な数の調査シートの存在

A 町のボランティアセンターが担当する地域では、津波による倒壊を免れた戸宅で生活をしている世帯が約 2,000 世帯、仮設住宅で生活をしている世帯が約 2,000 世帯あり、定期的な巡回が必要だと判断された世帯はおよそ半数の約 2,000 世帯にものぼる。現在は 3 チーム体制で 1 日あたり約 100 世帯訪問調査している。これに対して、センターにて Excel への入力を担当しているスタッフは二人のみである。著者らの協議会での書き起こし作業経験に基づく、一人が 1 時間書き起こし作業に集中しても 7~9 枚程度の調査シートを書き起こすのが精一杯であった。従って、1 日に 3 時間書き起こし作業だけに専念できたとしても、二人で 1 日に書き起こすことが可能な調査シートの枚数は高々 50 枚程度であり、書き起こすことのできない調査シートが毎日バインダーに綴じられるだけの状況になっていた。

## 2.4 クラウドソーシングを活用した書き起こし支援システムの開発

前節で述べた現場での被災者のケアにおける課題を解決するために、クラウドソーシングを活用した調査シートの書き起こし支援システムを開発し、A 町の社会福祉協議会に無償提供した。クラウドソーシングを活用することで、遠隔地にいながら現地スタッフに代わって、複数人による調査シートの書き起こしを支援できるようになり、2.3.2 節で述べた問題点の解決に繋がると考えた。2011 年 12 月 5 日現在、約 50 名の

ボランティアに参画いただき、1000 枚以上の調査シートの書き起こしを行っている。

図 2 にシステムの全体概要図を示す。本システムは、当社の提供する IBM Business Smart Cloud[4]上に配備されたサーバー、ボランティアセンターで稼動するクライアント、書き起こし作業を支援するための Firefox プラグインからなる。以下に、書き起こし支援システムを使用する、ボランティアセンタースタッフ、システム管理者、ボランティアの作業内容をまとめる。

### 2.4.1 ボランティアセンタースタッフの作業

ボランティアセンターのスタッフは、以下の作業を行うだけでよい。

- 管理されている調査シートにユニークな管理番号を付与し、順次スキャンして画像を作成する。
- 従来通りのやり方で調査シートの一部を Excel ファイルに直接書き起こす。
- クライアントシステムを操作して、調査シートのスキャン画像をサーバーに送り、同時にサーバーと Excel ファイルの同期をとる。

### 2.4.2 システム管理者の作業

サーバーのユーザーインターフェースは管理者ビューとボランティアビューの 2 つから構成されており、ボランティアセンターのクライアントからスキャン画像が届くと、まずは管理者ビューに登録される。

現在のシステムでは、2.3.1 節で述べたように、シートの変遷に伴い様々な種類の調査シートを扱う必要がある。そのため、システム管理者は管理者ビューを



生活支援シート(世帯の基本情報)

在宅者: 仮設入居 No. 2-57 訪問後日: 23年10月2日( )

|             |                                                                                                                                                                                                           |    |       |       |                  |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-------|-------|------------------|
| フリガナ        | サイトウ マサトシ                                                                                                                                                                                                 | 性別 | 男     | 生年月日  | 明・大田 19年(西暦/95年) |
| 世帯主名        | 斉藤 政俊                                                                                                                                                                                                     | 年齢 | 41(男) | 電話番号1 | 0120-345-6789    |
| 地区          | 担当民生委員                                                                                                                                                                                                    | 職業 | なし    | 電話番号2 |                  |
| 住所<br>(前住所) |                                                                                                                                                                                                           |    |       |       |                  |
| 家族構成        | 1 ひとり暮らし(別居の家族なし) 2 ひとり暮らし(別居の家族あり)<br>3 高齢者・障害者のみの世帯(夫婦等) 4 高齢者・障害者がいる世帯<br>5 母子・父子世帯 6 乳幼児がいる世帯<br>7 その他 世帯構成 ( ) ペット いる・いない ( )<br>注)世帯員が複数いる場合、下掲の世帯各欄に記入(世帯主には※をつける。上記の本人は記入不要。個別シートのある方に捺印の横に○を付ける) |    |       |       |                  |
| 家族・親族の被災状況  | 1 全員無事 2 死傷者あり ( ) 3 連絡が取れない ( )<br>4 家族の被災状況 1 全壊 2 半壊 3 大規模半壊 4 一部損壊 5 被害なし 6 被害なし 7 被災証明 あり・なし                                                                                                         |    |       |       |                  |
| 仮設入居前の避難状況  | 1 避難所 2 福祉避難所<br>3 緊急入院・入所等 4 その他                                                                                                                                                                         |    |       |       |                  |

生活再建に向けた各種支援申請状況及び経済状況

申請済 申請中 未申請 申請済 申請中 申請済 申請中 申請済 申請中 申請済 申請中

SDSS 書き起こし 完了 保存 キャンセル 再読み込み 一覧

管理番号 (1) さわ-12346

世帯種別 (1) 仮設

初回面接日 (1) 2011/10/25

フリガナ (1) サイトウ マサトシ

世帯主名 (1) 斉藤 政俊

性別 (1) 男性

生年月日 (1)

年齢 (1)

職業 (1)

地区 (1)

担当民生委員 (1)

仮設入居先

図3 Firefox プラグインによるボランティアビュー

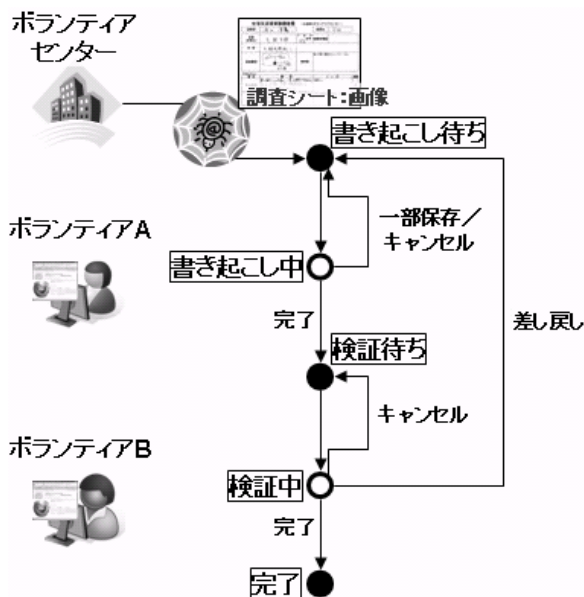


図4 書き起こし作業の処理フロー

活用して、事前にシートごとに記載されている情報の意味付けを行うことで、書き起こされたデータを有効活用できるようにする。また、管理者はこの作業以外に、管理番号の書き起こし、Excel に書き起こされている既存データとのマッチング、スキャン画像の歪み修正なども行う。管理者ビューでの作業を終えると、スキャン画像はボランティアビューに登録され、ボランティアに通知される。

## 2.4.3 ボランティアの作業

ボランティアは図4に示す作業フローにしたがって、事前にインストールしておいた Firefox のプラグイン(図3を参照)を利用して書き起こし作業および検証作業を行う。Firefox プラグインでは、(1)による画像の拡大/縮小、(2)による書き起こし対象項目のフォーカス、(5)によるボランティアの作業履歴の確認をサポートしており、画面左側の画像を見ながら(4)に示す各項目に書き起こしていく。作業の完了/キャンセルは(3)のボタンを通じて行う。

ボランティアビューに新しいスキャン画像が届くと、その画像の処理状態が「書き起こし待ち」として登録される。あるボランティア A が「書き起こし待ち」状態の画像を見つけ、書き起こしのため画像を選択すると、画像の処理状態が「書き起こし中」に変化する。これにより他のボランティアが同一の画像に対する書き起こし作業ができないよう排他制御を実現している。ボランティア A が何らかの理由により書き起こし作業を中断する場合には、それまでの作業内容を保存して終了することで、他のボランティアに作業を引き継ぐことが可能である。ボランティア A が書き起こしを終え、完了ボタンを選択すると、画像の処理状態は「検証待ち」へと変化する。本システムでは、第三者の目で書き起こし内容を確認することを前提として、一人のボランティアが同一画像に対する書き起こし作業と検証作業を続けて行えない。

別のボランティア B が「検証待ち」状態の画像を

見つけ、書き起こされたデータの検証のため画像を選択すると、画像の処理状態が「検証中」へと変化する。書き起こし時と同様に、他のボランティアが同一の画像に対する検証作業ができないよう排他制御を実現している。ボランティア B は書き起こされたデータに修正が必要だと判断した場合には、その理由も付して差し戻すことができ、この場合画像の処理状態は「書き起こし待ち」へと戻される。また、ボランティア B は検証作業を途中でキャンセルすることもでき、この場合画像の処理状態は「検証待ち」へと戻る。ボランティア B が検証作業で書き起こしデータに問題のないことを確認し、完了ボタンを選択すると、画像の処理状態が「完了」となり、全ての書き起こし作業および検証作業が完了したこととなる。

### 3. システム管理者作業を軽減する試み

現在のシステムでは、システム管理者が 2.4.2 節で述べた管理者作業を全て担うため、負担が大きくなっている。そこで、これらの管理者作業に既存技術やクラウドソーシングを活用することで、管理者の負担を軽減できないかを検討した。

1. スキャン画像の歪み修正作業：事前に登録しておいた歪みのない教師画像を利用して与えられた画像の歪み具合を検出する既存技術[5]を組み合わせて、画像の歪みを自動修正することで、管理者作業から除外可能である。
2. 管理番号の書き起こし作業：現在のシステムでは、事前に管理者が行うExcelの既存データとのマッチング作業に必要であるという理由で管理者作業になっているが、調査シートに記載されている他の手書き情報の書き起こしと実質的に何ら変わりはないため、クラウドソーシングすることで管理者作業から除外可能である。
3. Excelの既存データとのマッチング作業：Excel上で書き起こしの間違いがあったり、調査シート上は旧字体で記載されているにも関わらずExcel上は新字体で電子化されていたりと、マッチング作業時に実際にExcelを眺める必要性が生じたため、現在のシステムでは管理者作業になっている。しかし、作業の本質は該当データを検索することであるため、類似文字列尺度に関する研究[6]等を有効活用して検索精度をあげることでクラウドソーシングでき、管理者作業から除外可能である。
4. 調査シートに記載されている情報の意味付け作業：調査シート上で記載されている各情報がど

ういう意味をもつのかを判断するには、一般的にドメイン知識が必要となる。そのため、あまりドメイン知識のない作業員への単純なクラウドソーシングだけでは作業効率の向上が見込みづらいため、管理者作業から除外できるかどうか現時点で定かではない。

本章では、4 の作業にクラウドソーシングおよび自然言語処理を活用することで管理者作業の軽減につながるかどうかを、管理者作業量および書き起こし結果、書き起こされたデータの検索精度の点から考察する。

#### 3.1 各作業員に割り振るタスクの定義

調査シートに記載されている各情報は罫線で区切られた領域内に記載されているため、帳票から罫線を抽出する既存の画像処理技術[7]を用いて、まず領域を抽出し、情報ごとに分割する。本章における作業員のタスクは以降、情報ごとに分割された領域を対象として行うものとする。

調査シートの各情報はその情報の意味を key とした単純な key-value ペアであると考えことができ、調査シートはその key-value ペアの集合である定義できる。例えば図 3 における生活支援シートでは、フリガナ情報は (key, value) = (フリガナ, サイトウマサトシ) で表現でき、調査シート全体は {(居住形態, 仮設入居), (初回面接日, 2011/10/25), (フリガナ, サイトウマサトシ), (世帯主名, 斉藤政俊), …} という集合であるとみなせる。各作業員はタスクにおいて、これらの key-value ペアを順次構築していく作業を担う。作業フローを図 5 に示す。

#### 3.2 サブタスク 1: シートに記載されている情報の書き起こし

作業員は分割された領域画像を見ながら、記載されている情報を単純に書き起こしていく。これは key-value ペアにおける value を構築する作業にあたり、既存の書き起こしシステムで行っていた作業とほぼ同様の作業となる。作業の正確性を保障するため、複数の人に同じ作業を割り振ることとする。また、現在のシステムと同様に検証のフェーズも設ける。

#### 3.3 サブタスク 2: シートに記載されている情報の意味付け手順

作業員は分割された領域画像および書き起こされたデータを見ながら、そのデータに意味付けを行う。これは key-value ペアにおける key を構築する作業にあたる。[8]で述べられているように、クラウドソーシングで得られた意味付け結果を自然言語解析の教師データとして利用することは有効である。本サブタスクでは、作業員によって付けられた情報への意味付けを

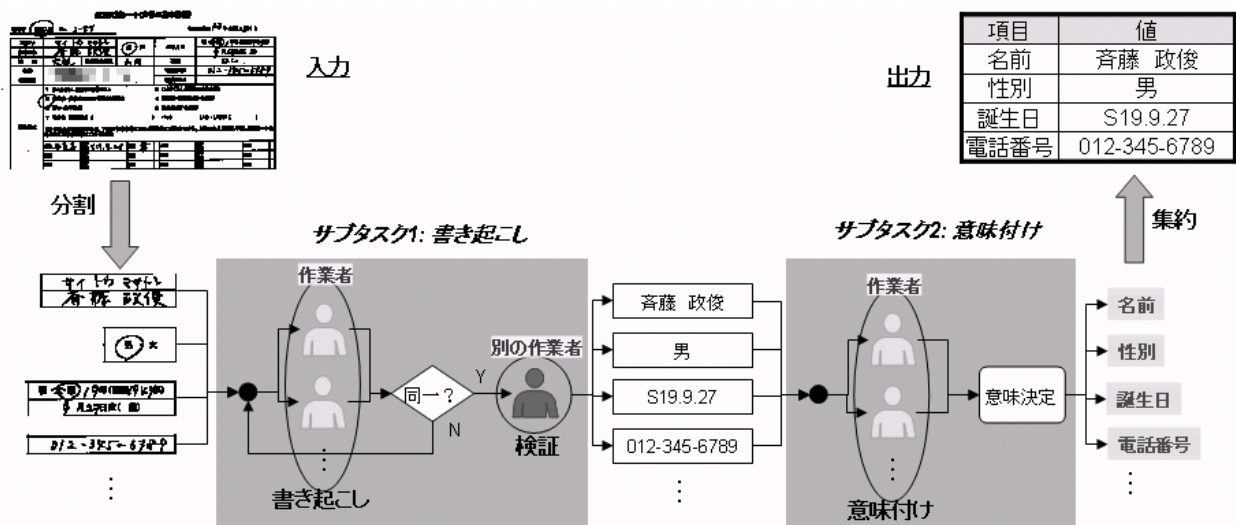


図5 作業者の処理フロー

データのカテゴリ推定を行う自然言語解析の既存技術[9]への教師データとして活用することで自動判別の精度を高めることを狙う。また、自動判別の結果を作業者にも提示することで、相乗効果による意味付けの正確性の向上を図る。

### 3.4 評価

発表当日までに実際の調査シートに対して試し、管理者作業量および書き起こし結果・検索精度の点から現在のシステムと比較する。クラウドソーシングの有効性を議論する上では、意味付けを自然言語解析だけで行う場合と比較して、精度やコストが大幅に向上していなければならない[10]。

## 4. おわりに

本発表では、管理者の作業負担を軽減し、クラウドソーシングをより使いやすいものとするため、調査シートの書き起こし支援システムを対象に一般的にはドメイン知識を有する専門家の助けが必要な管理者作業の一部をもクラウドソーシングすることを試みた。発表当日には、管理者作業量および書き起こし結果、書き起こされたデータの検索精度の観点から考察した結果も合わせて示す。

## 参考文献

- 1) 東日本大震災の被災者数, <http://www2.ttcn.ne.jp/honkawa/4362.html>. (2011年12月5日現在)
- 2) 市口恒雄, 災害時にもロバストで有効な情報伝達手段, 科学技術政策研究所 科学技術動向, 2011年6月号, pp.8-20, 2011.
- 3) 近藤範子, 災害のストレスによる慢性期・復旧復興期の心身の健康問題と心のケア, 日本看護

協会出版会 Nursing Today, Vol.26, No.4, pp.18-22, 2011.

- 4) IBM Smart Business Cloud - Enterprise, <http://www-935.ibm.com/services/jp/igs/cloud-development/>.
- 5) G. Rosman, A. Tzadok, D. Tal, A New Physically Motivated Warping Model for Form Drop-Out, ICDAR2007, pp.774-778, 2007.
- 6) N. Okazaki, J. Tsujii, Simple and Efficient Algorithm for Approximate Dictionary Matching, COLING 2010, pp.851-859, 2010.
- 7) H. Tanaka, H. Takebe, Y. Hotta, Robust Cell Extraction Method for Form Documents Based on Intersection Searching and Global Optimization, ICDAR 2011, pp.354-358, 2011.
- 8) C. Akkaya, A. Conrad, J. Wiebe, R. Mihalcea, Amazon mechanical turk for subjectivity word sense disambiguation, NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, pp.195-203, 2010.
- 9) ローネン・フェルドマン, 辻井潤一監訳, IBM 東京基礎研究所テキストマイニングハンドブック 翻訳チーム訳, テキストマイニングハンドブック, 東京電機大学出版局, 2010.
- 10) E. Adar, Why I hate Mechanical Turk Research (and Workshops). ACM CHI Workshop on Crowdsourcing and Human Computation, 2011.