

## 東日本大震災と情報、インターネット、Google

### 第7回：パーソンファインダー、東日本大震災での進化（2）



「避難所に手書きの名簿があるはずなので、それを撮影して投稿してもらおう」 Google が募集をかけたところ、大量の写真が集まった。さらに 5000 人ものボランティアが協力して写真を読み取り、パーソンファインダーにデータを登録していった。

2012 年 4 月 13 日掲載

### 名簿写真から名前を読み取ってデータを手入力

3 月 14 日（月）の朝、東京オフィスに出勤した Google マップのプログラムマネージャ、村上陽祐は、Picasa ウェブ アルバムを見て驚いた。午前 2 時にスタートしたばかりの「避難所名簿共有サービス」に、すでに大量の写真データが投稿されていたからだ。

ただし、避難所名簿共有サービス単独だとそれほど利便性は高くない。確かに避難所名を頼りに 1 枚 1 枚写真を見ていけば、知人や家族の安否を確認できることもあるだろうが、これにはかなりの手間がかかる。やはり人名などのデータから検索できるようにするのが望ましい。Google は自社のミッションを「世界のあらゆる情報を整理して世界中の人がアクセスできるようにすること」としており、この思想は社員の中にも浸透している。

では、画像データとして送られてきた名簿をどうやって検索できるようにするか。

Google は高度な画像処理技術を自社開発している。例えば、Google ブックスでは、ブックスキャナが書籍のページをめくりながら自動的に撮影、画像データから文字を抽出してデジタルデータ化する。このように、画像データ内の文字を認識してテキストデータに変換する仕組みを OCR (Optical Character Reader / Recognition) という。OCR 技術は年々進歩しており、手書き文字でもかなりの認識率を示すようになってきているが、まだ人間には及ばない。特に、混乱した避難所の名簿の文字は、走り書き、殴り書きだ。米国のチームが OCR の技術を使って作業効率を上げられないか検証をしたが、実現性が低いということがわかった。

Google 社内では避難所名簿をデジタルデータ化するためのさまざまな方策が検討されたが、家族・知人の安否を求める人々に一刻も早く情報を届けるため、一番原始的な方法を採用することにした。つまり、人間が 1 枚 1 枚の写真を見て、書かれている名前その他の情報を読み取ってパーソンファインダーに入力するので

ある。

村上是、どうしたら効率よくテキスト化の作業を行えるか思案した。新たにシステムを開発するとなると時間がかかりすぎるため、Google ドキュメントのスプレッドシートを社員で共有し、進捗状況を管理する。Picasa ウェブ アルバムにアップロードされた写真には、それぞれ異なる URL (インターネット上の住所) が割り振られているので、これをスプレッドシートにコピー。Picasa ウェブ アルバムに投稿された写真にはコメントを付けられるようになっているので、ここに「作業開始」「完了」と記入して、作業の重複を防ぐ。そして、スプレッドシートに入力したデータをパーソンファインダーへと転記していくようにすればよさそうだ。

村上是ふだんいっしょに仕事をしているチームのメンバーに声を掛け、社内メーリングリストでもボランティアを呼びかけた。50 人ほどのメンバーが集まったので、図版入りの手順書を作って作業内容を説明し、14 時ころには作業をスタートさせた。

15 日 (火) になると、社内ボランティアの数は 85 人にまで増えていた。作業している人間が普段担当している業務はさまざまで、マーケティング担当やセールス担当もいれば、Google 日本語入力やウェブブラウザの Chrome を開発しているエンジニアもいた。ふだんは難解なプログラムコードと格闘しているエンジニアも、写りの悪い名簿の写真から何とか名前を読み取るうと奮闘していた。また、基礎技術を担当しているエンジニアは、作業を効率化するための Chrome 拡張機能を開発し、15 日の午後に社内でも公開した (23 日 (水) には一般向けにも公開された)。この拡張機能は、Picasa ウェブ アルバムのコメント欄から情報をコピーしてパーソンファインダーに転記してくれるというものである。海外オフィスからも、日本語の読み書きができる社員が何人か作業に参加し、タイムゾーンの異なる東京オフィスのサポートに当たった。

## テキスト起こし作業を外部のボランティアに依頼

だが、実際に作業を進めていくうちに、写真のテキスト化には予想以上に手間がかかることがわかってきた。1 枚の写真に 1 人の名前だけが写っていることもあれば、100 人のこともある。文字が不鮮明な写真や、読み方のわからない名前や住所も多いため、1 人が 1 枚の写真をパーソンファインダーに入力するのに 1 時間かかることも珍しくない。

15 日までに Picasa ウェブ アルバムに投稿された写真の枚数は 1800 枚にも上ったが、社内でもパーソンファインダーに入力できたのは 1 / 3 の 600 枚程度にすぎなかった。しかも、写真投稿はさらに増える気配を見せていた。

当初は写真の読み取りからパーソンファインダーへの入力まで社内ボランティアで行う予定だったが、作業に無理が出てきたのは明らかだった。一部では名簿のテキスト起こしを始めているユーザーもいた。ならば、外部のボランティアに対して正式に協力を仰ぐべきか? 悪意あるユーザーがイタズラするということはないか?

**避難所名簿共有サービス**



避難所にある名簿の写真を撮影して、メールのタイトルに避難所の名前を入れて [tsukasa.sugi@google@picasaweb.com](mailto:tsukasa.sugi@google@picasaweb.com)宛に送信してください。送信された写真は Picasa web album にアップされ、それらの写真は <http://goo.gl/ganbare> から見るすることができます。詳細はこちら。

[名簿画像のアップロード、画像からデータを入力するボランティアはこちら](#)

クライシスレスポンスサイトのトップに掲載された、避難所名簿共有サービスへの投稿を呼びかける案内文。被災地でのリアルな雰囲気がこの写真から伝わってくる。

懸念はあったが、ユーザーの善意を信じて作業のスピードを優先することになった。

告知はまず Picasa ウェブ アルバムで行われた。15 日の早朝、村上はアルバムに「テキスト起こしのお願い」と題した「画像」をアップし、名簿写真のコメント欄に「姓 名（その他の情報）」という書式でテキストを入れてくれるよう、ボランティアにお願いした。

## データ登録プロセスが自発的に作られていった

Picasa ウェブ アルバムの避難所名簿共有サービスには、数多くのボランティアが集まってきており、写真のコメント欄を使ってのディスカッションが自然と始まった。

尼崎市在住のソーシャルメディアクリエイター、OYAJI さん（ハンドル名）は早い段階からパーソンファインダーや避難所名簿共有サービスの存在を知り、Twitter で使い方などの情報発信を行っていた。避難所名簿をテキスト起こしする必要があるのではないかと考えていたところに、Google による「テキスト起こしのお願い」画像がアップロードされたのを見て、Picasa コメント欄でのディスカッションに参加し始めた。OYAJI さんや、うずまいさん（ハンドル名）らボランティアは、コメント欄にテキスト起こしをする際の書式や手順が統一されていないとパーソンファインダーへの入力作業が混乱することを懸念し、話し合いながら詳細を固めていった。個人情報保護の観点から「番地は写真に記載があってもテキスト化しない、パーソンファインダーに登録しない」という方針も、社外ボランティアたちが率先して決めていったという。

また、ウェブサイト制作者の tenkao さん（ハンドル名）は、未作業のファイルを探すためのサービス「Google 避難所名簿共有サービス アルバム・写真リスト」を立ち上げていた。このようなサービスやツールは、ボランティアの作業を効率化するのに大いに役立った。

16 日（水）になると、パーソンファインダーへの入力作業も Google の社内ボランティアだけでは追いつかなくなり、社外ボランティアが本格的に作業に参加するようになった。社外



Picasa ウェブ アルバム上でのテキスト化作業の様子。写真のコメント欄を使って、ボランティア同士が意見交換しながら作業を進めていった（図は作業プロセスが固まってきた 3 月 19 日時点のもの）。

ボランティアは、ディスカッションを通じて作業プロセスを固め、17日(木)0時にはうずまいさんが「Person Finder テキスト化 & 登録まとめ Wiki」(<http://www45.atwiki.jp/ganbare-tohoku/pages/1.html>)を作成して情報を集約した。上記サイトは現在も残っているが、うずまいさんやつくば市の Oblivion さん(ハンドル名)らによって、未作業の写真データを探す手順から、作業の開始/終了の宣言、テキストの書式、パーソンファインダーへの登録手順、登録したデータのダブルチェック、トリプルチェックまで、図版入りで丁寧に解説されている。さらに、外国語名や難読地名の手がかりとなる情報まで網羅

した労作である。パーソンファインダーが改良されて仕様が変更されるたびにデータ登録のルールも変える必要があったが、それらもまとめ Wiki に逐一反映されていった。

17日の午前8時40分には Google 公式ブログ 日本版にも避難所名簿のパーソンファインダーへの登録作業について協力を求めるコメントが掲載された。しかし、IT に詳しくない人でも何をしたらいいのかわかるようにするには、図版を使うなど見せ方も工夫する必要がある。そこで、Google の UX (ユーザー体験、つまりユーザーとどのように接すればいいか) を担当している石塚尚之や Web マスター (Web ページ制作) の小山結末らがボランティア向け説明ページの制作に当たった。

翌18日(金)の14時までには、3360名の社外ボランティアが登録作業に参加し、送られてきた名簿画像の90%以上がパーソンファインダーへと無事登録された。

社外ボランティアが本格的に協力し始めてからわずか2、3日で、Google 社内だけでは処理しきれなかった名簿データをさばくことができたことになる。

その後も社外ボランティアの活動は続いた。大阪のふありんさん(ハンドル名)は、登録済みとされた90%の中にも、文字が判読できないなどの理由で、事実上埋もれたままになっているデータがあることに気づき、再チェックのプロセスを整えていった。震災から2週間後に参加した YokoS さん(ハンドル名)によれば、この頃になると文字がはっきり読み取れる写真についてはほぼ作業が終わり、残っていたのは解像度が低くて読みづらいもの、崩し字などであった。編集の経験があった YokoS さんは、こうした写真データの読み取りや、読み違いのチェックなどを手がけた。また、ボランティアの間では校正を行って誤字や読み違いを正し、データの精度を上げていこうという声上がり、4段階ほどの校正プロセスが固まっていった。読み取りにくい文字についても、「◎◎かもしれない」というコメントを付けて複数の作業者を経ることで判別できたことも多いという。

29日(火)、Google は、参加ボランティアの合計は5000人に達し、処理した画像は1万枚以上、登録



3月17日(木)14:00時点でのパーソンファインダーへの登録状況。この時点では投稿された写真のうち、まだ1/3程度のデータしか登録されていなかった。

したデータ件数は 14 万件以上になったと発表した。

6 月に避難所名簿共有サービスが閉鎖されるまで、社外ボランティア達は作業を続けた。

ちなみに、名簿のテキスト起こしボランティア作業は、上記のまとめ Wiki だけで行われたわけではない。mixi のコミュニティ上で声を掛け合いながらテキスト起こしを進めていた人、個人ブログ上で独自に作成した名簿を投稿していた人も多かった。OYAJI さんによれば、ブログや mixi、そしてまとめ Wiki 上の情報を突き合わせて初めて、被災者の身元の確証が得られることもあったという。

## プラットフォームを提供することで、人々がつながっていく

社外ボランティアが参加したからこそ、14 万件もの安否情報が驚くほど短期間に入力された。しかし実を言えば、社外ボランティアと Google は必ずしも密に連携できていたわけではない。

例えば、当初パーソンファインダーには、名前での検索しか行えず、100 件以上の結果は表示できないという仕様になっていた。社外ボランティアのエンジニアは、Google グループでパーソンファインダー改善の要望を投稿していたが、これに対する Google からのレスポンスはほとんどなかったため、ストレスを感じていた人もいたようだ。

実際のところ、パーソンファインダーの改良については、ソフトウェアエンジニアの川口良や高橋周平らが精力的に取り組んでいた。17 日（木）にはかな・ローマ字での曖昧検索が追加され、3 月末には住所と名前の組み合わせ検索も可能になっている。このほか、ユーザーが未着手の写真を簡単に探せる方法を用意し、社外ボランティアが進めている作業と協業できるようにするなど細かく気を配った。だが、Google 側の内情が説明されなかったこと、パーソンファインダーの改善状況や予定についてアナウンスがなされなかったことなど、コミュニケーション不足があった面は否めない。

データ登録作業のプロセスで、Google と社外ボランティアの間に齟齬が生じたこともあった。

例を挙げると、社外ボランティアは、同一避難所で撮影された写真が複数枚ある場合、それらを総合的に判断してデータを登録するというプロセスを作っていた。別々の写真になっていても家族や親族と判断できれば、その旨をパーソンファインダーの情報欄に付記して、消息を探しやすく工夫していた。ところが Google 側で、こうしたプロセスを踏まずに朝日新聞に掲載された名簿からの情報をスピード優先で入力したため、大量の重複登録が生じたケースがあった。スピードと精度、どちらを優先するかの是非を判断するのは難しい。ただ、(同じ名前のデータが 100 件以上あった場合) 100 件以降の検索結果が表示されないというパーソンファインダーの仕様のため、重複したデータを探すのに手間がかかるようになってしまったようだ。

また、当初 Google 側は「避難者名簿以外の写真は、パーソンファインダーへの入力は避ける」という方針を示していたが、途中からこれらの写真がテキスト起こし対象のフォルダに入れられるようになるなど混乱が生じた。こうしたデータの扱いも、ボランティア同士で相談し対応を決めていった。

コミュニケーションやプロセスの面では課題もあったとはいえ、パーソンファインダーは IT が人々をつなぐことを示す好例となった。知名度の高い組織がプラットフォームさえ提供すれば、人々は自発的に支援活動を進めていくことができる。シニアエンジニアリング マネージャーの賀沢秀人は、今回の事例を「預ける」(情報を公開・共有する)、「見つける」(情報を活用する人が現れる)、「広める」(ネットを使って拡大する)という 3 つの要素がうまくかみ合った結果だと分析する。

読者のひさおさんからは、「写真からテキストへ、そしてパーソンファインダーへの入力が、一般ユーザーに託されるまで時間が掛かりました。もっと早く、ユーザーを信頼してほしいです。それ以外は素晴らしい

く、1箇所にもとめる重要性を感じさせました」というコメントをいただいた。

また、OYAJIさんは「今回の名簿テキスト化作業では、Google社内グループ、まとめWikiのグループ、mixi上のコミュニティ、ブログのグループ、個人、マスコミなどが別個に動いていました。こうしたグループが相互に協議・情報交換できる場を提供していただけたら、Google側との齟齬も少なくなり、各グループへの統一ルールへの浸透も図れ、情報作成・提供のスピードはさらに速くなったのではないのでしょうか」と語る。

自治体やマスコミ、そして一般ユーザーの間に安否情報の提供に関するノウハウが浸透してくれば、今後の災害が起こった時にもより早くより正確に情報を伝えることが可能になるはずだ。安否情報を伝えるための手順、外部ボランティアの募り方、ボランティアが活動しやすいコラボレーションサービス、コミュニケーション窓口の設置など、今回の経験を元に考えるべき対策は多岐にわたり、多くの人々の知恵が必要とされている。特に、被災地のために何か手助けをしたいと思いつつ、結局何もできずに罪悪感を抱いた人は少なくないのではないか。こうした人々の善意を活用できるプラットフォームをいかにして構築するかは大きなチャレンジだろう。

取材、執筆、編集：林信行 / 山路達也