

平成 29 年（2017 年）5 月 23 日

「サンプル名刺データ」を研究用データセットとして無償提供開始 国立情報学研究所とクラウド名刺管理の Sansan 株式会社が提携

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所（NII、所長：喜連川 優、東京都千代田区）と Sansan 株式会社（Sansan、代表取締役社長：寺田 親弘、東京都渋谷区）は研究コミュニティへの研究用データ提供で提携^(*)し、NII が 5 月 23 日より Sansan の所有する名刺情報のサンプルデータを「Sansan データセット」（<http://www.nii.ac.jp/dsc/idr/sansan/sansan.html>）として研究コミュニティに無償提供します。

提供するデータは、Sansan が昨年実施したデータサイエンティスト向け分析コンテスト^(**)で使用された、サンプル名刺をスキャンした画像データ 3841 枚分です。最大 9 種類の項目（会社名、名前、役職、郵便番号を含む住所、電話番号、FAX 番号、携帯電話番号、電子メールアドレス、web サイトの URL）の情報を含んでいます。



サンプル名刺データの画像の例

本データは Sansan が自社で作成した架空の名刺であり、特定の企業や個人につながる情報や人物を特定できる個人情報は一切含まれておりません。「Sansan データセット」は、NII のデータセット共同利用研究開発センター（センター長：コンテンツ科学研究系教授 大山 敬三）の情報学研究データリポジトリ（IDR）を通じて提供されます。

【サンプル名刺データ提供の背景】

NII は平成 22 年（2010 年）に IDR を設置し、様々な企業や機関が保有する各種のデータセットを受け入れて研究コミュニティに提供を行ってきました。平成 27 年（2015 年）4 月には、情報学研究に有用なデータセットを整備して研究者に提供するとともにデータセットの構築とその活用基盤に関する研究開発を行うデータセット共同利用研究開発センターを新設し、研究コミュニティへのビッグデータの提供をさらに強化するとともに、大規模な実データと最先端情報技術を活用したデータサイエンス研究の加速に取り組んでいます。

一方、Sansan では、通常のドキュメントに比べて非定型でフォントやフォントサイズ、レイアウトなどのフォーマットが決まっていない名刺を年間数億枚も正確にデータ化しなければなりません。その

ため、画像処理、機械学習のスペシャリストやデータサイエンティストがテクノロジーを進化させるような専門性の高い研究に取り組み、独自開発した名刺データ化システムと人カオペレーションを組み合わせたデータベースを構築しています。

個人情報である名刺はデータソースが限られるため、大学や研究機関での名刺の画像解析の研究は活発化しにくいのが現状です。こうした背景から、Sansan では昨年にデータサイエンティスト向け分析コンテスト^(*)を開催するなど、オープンイノベーションを積極的に進めてきました。今回、NII を通じてサンプル名刺データを広く研究コミュニティに提供することにより、画像処理や機械学習の研究の発展に貢献できることを期待しています。NII では、今後もデータセットの拡充とデータセット共同利用の深化のための研究開発に取り組み、より多くの研究者が自由な発想で研究に利用できるようにするためのシステムプラットフォームの構築を進めてまいります。

以上

《国立情報学研究所（NII）について》

NII は情報学という新しい学術分野での「未来価値創成」を使命とする国内唯一の学術総合研究所です。情報学における基礎論から人工知能やビッグデータ、Internet of Things (IoT)、情報セキュリティといった最先端のテーマまでの幅広い研究分野において、長期的な視点に立つ基礎研究、ならびに、社会課題の解決を目指した実践的な研究を推進しています。また、大学共同利用機関として、学術情報ネットワーク（SINET5）をはじめ、学術コミュニティ全体の研究や教育活動に不可欠な学術情報基盤の構築・運用に取り組むとともに、学術コンテンツやサービスプラットフォームの提供などの事業を展開・発展させています。

《Sansan 株式会社について》

Sansan は平成 19 年（2007 年）の創業より、「ビジネスの出会いを資産に変え、働き方を革新する」をミッションに掲げ、「出会い」の価値を最大化するクラウド名刺管理サービスを開発・提供しています。「Sansan」は「名刺を企業の資産に変える」をコンセプトに、社内に眠る名刺をデータ化して人と人のつながりを可視化し、シェアできるクラウド名刺管理サービスで、導入社数は 5500 社を超えています。平成 24 年（2012 年）からは、ソーシャルの仕組みを取り入れて名刺を新たなビジネスネットワークに変える、名刺アプリ「Eight」も提供しています。登録ユーザー数は 150 万人を超えています。

〈メディアの皆様からのお問い合わせ先〉

大学共同利用機関法人 情報・システム研究機構

国立情報学研究所

総務部企画課 広報チーム

TEL:03-4212-2164 FAX:03-4212-2150

E-mail : media@nii.ac.jp

Sansan 株式会社

ブランドコミュニケーション部 広報担当

TEL:03-6316-3121 FAX:03-3409-3133

E-mail : pr@sansan.com

本件は NII と Sansan が共同で発表するものです。NII から文部科学記者会と科学記者クラブを通じて加盟メディアの皆様へ資料提供しているほか、NII、Sansan それぞれから関係各メディアの方々に個別に本リリースをお送りしています。重複して配信される場合がありますことをご了承お願いいたします。

(*) 「新たに提携」： 情報・システム研究機構と Sansan 株式会社が 5 月 22 日付で「データ提供及び利用に関する契約書」を締結。

(*) データサイエンティスト向け分析コンテスト： 「人工知能は名刺をどこまで解読できるのか」
(<https://jp.corp-sansan.com/lp/data-sientist-c.html>)