

2022年(令和4年)11月24日

「地球の歩き方」の利用者投稿旅行記データを 学術研究用に無償で提供開始

株式会社地球の歩き方(地球の歩き方、東京都品川区、代表取締役社長:新井 邦弘)と情報・システム研究機構 国立情報学研究所(NII、東京都千代田区、所長:喜連川 優)は学術研究分野でのデータ活用促進に向けて提携し、地球の歩き方が運営するウェブサイトの「旅スケ」サービスに投稿された旅行記のテキストデータを、研究用に「地球の歩き方旅行記データセット」として11月24日(木)より無償提供を開始しました。

このデータセットは、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)が運営する情報学研究データリポジトリ(IDR)を通じて提供されます。

地球の歩き方では、ガイドブック「地球の歩き方」を出版する傍ら、ウェブサイト「地球の歩き方 web」(<https://www.arukikata.co.jp/>)にて、旅に関する様々な情報発信や、旅行者同士の情報交換の場の提供をしています。

NIIは2015年4月にデータセット共同利用研究開発センターを設置し、様々な民間企業や研究機関が保有する各種のデータセットを受け入れて研究者に提供する「情報学研究データリポジトリ」(IDR)事業に取り組むことで、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。

このたび、地球の歩き方とNIIは学術研究分野向けのデータ提供で提携し、IDRを通じて、「地球の歩き方旅行記データセット」をアカデミア研究者向けに11月24日(木)から無償で提供することになりました(*1)。

今回提供を開始するデータは、「地球の歩き方 web」で2022年3月まで運営していた旅行記投稿サービス「旅スケ」に、サービス開始の2007年11月から2022年2月までに投稿された旅行記約14,000件とそれに紐づく旅スケジュール約9,500件のテキストで(*2)、具体的な例は以下の通りです。

旅行記

自由記述で入力されたテキスト

(国内旅行記約4,500件、海外旅行記約9,500件)

会津若松へ向かう磐越西線の接続を考慮して選んだやまびこ 203 号は E5 系での運転でした。何度も乗っている E2 系よりも座席が広く感じ、快適な移動でした。好天で磐越西線の車窓から磐梯山が良く見えたが、山頂だけが雲に覆われていました。

会津若松駅では手作り感のある赤べこのオブジェが出迎えてくれました。これから向かう伊佐須美神社の広告もありました。

伊佐須美神社へ向かう路線バス（永井野行）の出発点として若松駅前バスターミナルを利用しました。

...

会津若松駅から快速あいつ 4 号に乗り、郡山へ向かいました。会津若松は晴れ間がありましたが、山を上るにつれて雲が増えて行き、途中から雨が降り出しました。天気予報通りでしたが、今回の旅行は暖かい 2 日間で移動時間を除いて雨に降られることがなかったのはラッキーでした。

図 1 「旅行記」の例

旅スケジュール

旅行記に付随して任意で入力されたテキスト

(国内旅行記分約 3,100 件, 海外旅行記分約 6,400 件)

1 日目 2021 年 10 月 15 日(金)

...

08:19 - 08:29 郡山駅/東北新幹線

08:29 - 09:40 移動 (電車)

09:40 - 09:46 会津若松駅/磐越西線

09:46 - 10:20 若松駅前バスターミナル

10:20 - 11:07 移動 (バス)

...

2 日目 2021 年 10 月 16 日(土)

00:00 - 09:25 庄助の宿 瀧の湯

09:25 - 09:32 東山温泉入口 (瀧の湯前) バス停

09:32 - 09:36 移動 (バス)

...

図 2 「旅スケジュール」の例

旅行記テキストについては、奈良先端科学技術大学院大学 大内啓樹助教が研究代表者を務める JSPS 科研費基盤研究(B) 22H03648 の研究プロジェクト (<https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-22H03648/>) の協力により、自動解析による形態素情報 (単語の品詞等) と固有表現 (地名, 施設名等) をリストしたデータも併せて提供します。これらのデータは、自然言語処理分野における場所を参照する言語表現の分析や、場所に対する感情分析および評判分析などの研究への活用が見込まれ

ます。より高度な活用として、テキスト中の地名が実際に訪れた場所であるのか一般論としての言及であるのか周辺の文脈を読み解いたり、実際の地図上の位置と紐づけたりできるようになれば、旅行プランニングや観光地ナビゲーションといったアプリケーションへの応用も期待できます。

地球の歩き方と NII は、このデータセットの提供が、一層の学術研究の発展とオープンサイエンスの進展、及びデータサイエンスにおける教育と人材育成の強化に貢献できることを期待しています。

国立情報学研究所 情報学研究データリポジトリ 地球の歩き方旅行記データセット

<https://www.nii.ac.jp/dsc/idr/arukikata/>

【株式会社地球の歩き方について】

株式会社地球の歩き方（東京都品川区）は、1979 年に創刊した旅行ガイドブック「地球の歩き方シリーズ」を中心とする出版事業と、インバウンド・アウトバウンドの旅行マーケティング事業を行っております。ウェブメディア「地球の歩き方 web」は書籍に掲載できなかった最新の海外旅行情報や、各国在住の特派員や旅のスペシャリストが発信する旅行記事を展開し、海外旅行者を中心に旅の最新情報をお届けしています。

【旅スケ（旅スケジュール）とは】

「地球の歩き方 web」上で提供していた、旅のスケジュールや持ち物リストの作成、旅行記の公開等が可能な旅行者のコミュニケーションサービスです。（2022 年 3 月でサービスは終了）

【国立情報学研究所（NII）について】

国立情報学研究所（NII）（東京都千代田区）は、2000 年に設置された、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。

【情報学研究データリポジトリ（IDR）とは】

IDR は、NII のデータセット共同利用研究開発センターが運営するデータセットの共同利用事業です。IDR では各種のデータセットを民間企業や大学等研究者から受け入れて研究者に提供するためのサービスを行っています。

以上

〈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 E-mail : media@nii.ac.jp

株式会社地球の歩き方

コンテンツ事業部 web 運営室

TEL : 03-6431-1379 E-mail : web@arukikata.co.jp

-
- (*1) 地球の歩き方旅行記データセットの提供先は、原則として、地球の歩き方とNIIが定めた利用条件等に同意した大学や公的研究機関の研究室などに限りますが、利用条件等の範囲であれば、研究代表者の適切な管理のもとで、大学生・大学院生を含む研究者が自由な発想に基づく学術研究に利用することができます。また、データセットの利用状況や論文などの研究成果を定期的に報告いただき、NIIが研究成果の一覧を公開します。
- (*2) 地球の歩き方が投稿データを取得する際に、投稿者から著作権その他の著作権法上の権利の委譲を受けています。データに個人情報はありません。またサービス上では旅行記と合わせて写真の投稿が可能でしたが、提供データセットはテキストのみで写真は含まれません。