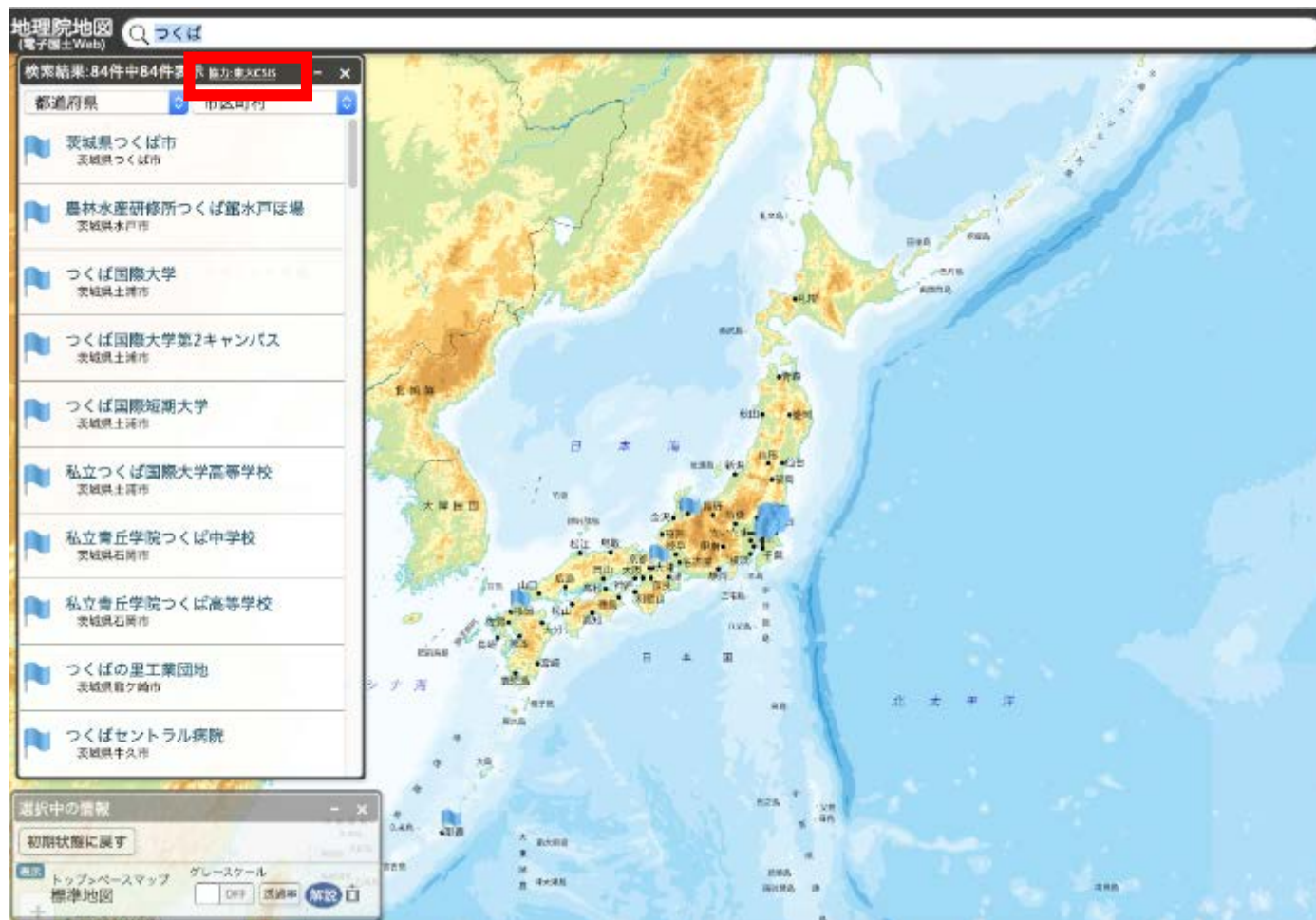


CSISジオコーディングサービスの活用 (国土地理院・地理院地図での利用例)

※逆ジオコーディングとしてもAPI上で稼働



<参考> GeoNLP: 文章から自動的に地名や住所を抽出する研究 & 地名解析辞書の提供システム (国立情報学研究所ほか)

GeoNLP 1.2.0 documentation » システム開発者向け資料 »

前へ | 次へ | 索引



GeoNLP

ショートカットリンク

- [GeoNLP サイトトップ](#)
- [ドキュメントトップ](#)
- [お知らせ](#)
- [辞書製作者向け資料](#)
- [システム開発者向け資料](#)
- [用語・フォーマット](#)

前のトピックへ

[getDictionaries](#)

次のトピックへ

[ダウンロード](#)

GeoNLP ソフトウェア

GeoNLP ソフトウェアは、自然言語テキストから地名を抽出するオープンソースの「ジオパスエンジン」です。たとえば「国立情報学研究所は千代田区にあります。」といったテキストから、「千代田区」を抽出し、その代表点の経緯度を取得することができます。

GeoNLP 公式ウェブサイト の WebAPI サービスも、本ソフトウェアを利用して実現しています。

使い方

本ソフトウェアには、コマンドラインで実行可能な [コマンドラインプログラム](#) と、ウェブサーバに設置して利用する [CGI プログラム](#) が含まれています。

CGI プログラムを利用すると、テキストデータを受け取って経緯度を付与する（その結果を地図にする）機能を含むウェブサービスを構築できます。コマンドラインプログラムを利用すると、地名や住所を含む大量のテキストデータに、一括処理で経緯度を付与することができます。

特徴

本ソフトウェアは、「[地名解析辞書](#)」に掲載されている [地名語](#) は確実に抽出することを基本方針としています。[地名解析辞書](#) を利用して形態素解析のレベルから制御し、異体字・新旧字などの表記揺れにも対応することで、抽出精度を向上させています。

[GeoNLP Data サイト](#) 上で公開されている任意の [地名解析辞書](#) をダウンロードして利用することができますので、従来の形態素解析器よりも多くの地名語や新しい地名語を抽出することができます。

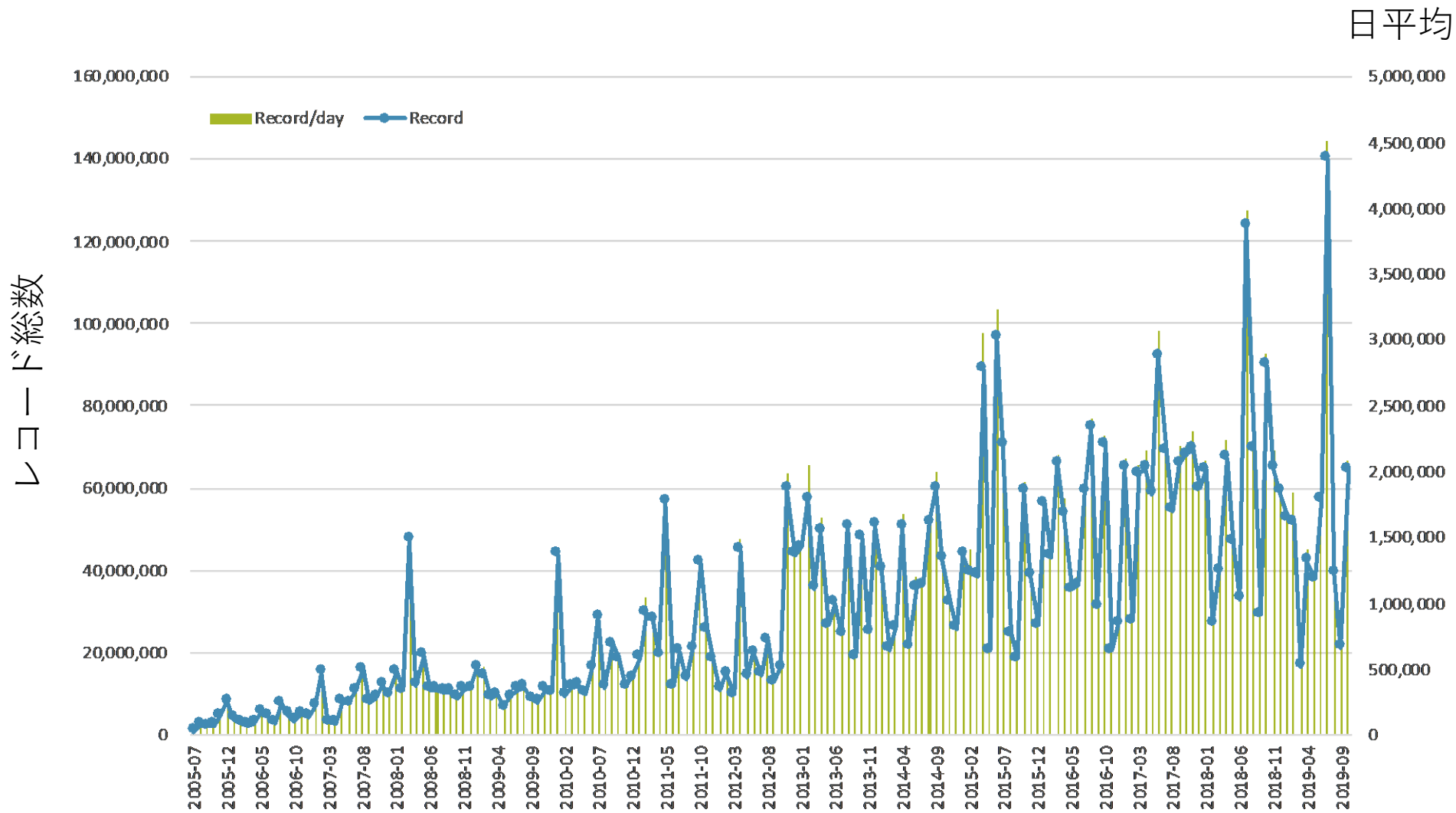
その反面、現状では「地名ではない語が地名として抽出されてしまう（過抽出）」ことがあります。用途に合わせて適切な [地名解析辞書](#) を選択したり、抽出結果を後処理でフィルタリングするといった方法で、十分実用的な活用が可能です。ソフトウェアの改良により、抽出精度も改善していく予定です。

Ver. 1.0.5 より、[東京大学空間情報科学研究センター](#)から公開されている[ジオコーダー DAMS](#) とリンクすることで、住所文字列を抽出して経緯度を算出できるようになりました。

以下、ローカルサーバ上に GeoNLP ソフトウェアをインストールし、利用する手順を説明します。

Source: <https://geonlp.ex.nii.ac.jp/>

CSISジオコーディングサービスによる 変換レコード数の推移

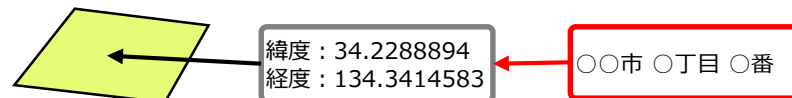


近年の利用実績 (2019年1月～11月時点)

- 利用ユーザー数
 - 約10万（1日平均約300ユーザー）
- 変換アドレス件数
 - 約7億446万レコード（1日平均約200万レコード）
- 「地理院地図」経由の検索数（上記とは別集計）
 - 約1,573万レコード（1日平均約47,000レコード）

「デジタル地図」を活用した農地情報における期待・課題（1）

1. 筆ポリゴンデータへ台帳で一般的に使われる住所データの付与（逆ジオコーディング or 行政データとのマッチング）



2. 筆ポリゴンと農地ポリゴンデータとの突合

（参考）農地ポリゴン（農地台帳）と筆ポリゴンについて

- 農地ポリゴン（農地台帳）は、農業委員会が登記簿ベースの情報を整理したもので、一方、筆ポリゴンは、50cm解像度の人工衛星画像等を用いて、統計部の職員等が人の手により、現況ベースで作成・整備したもの。
- したがって、農地ポリゴン（農地台帳）では、農地であっても実際には耕作されない場合等があることから、現況ベースの筆ポリゴンとは必ずしも一致しない。

農地ポリゴン（農地台帳）とは

- 農地法に基づき市町村農業委員会が農地台帳を整備。
- 全国農地ナビにて、農地の所在地（地番等）、面積、地目等の情報を公開。



水色:農地ポリゴン(農地台帳)、オレンジ色:統計部の筆ポリゴン

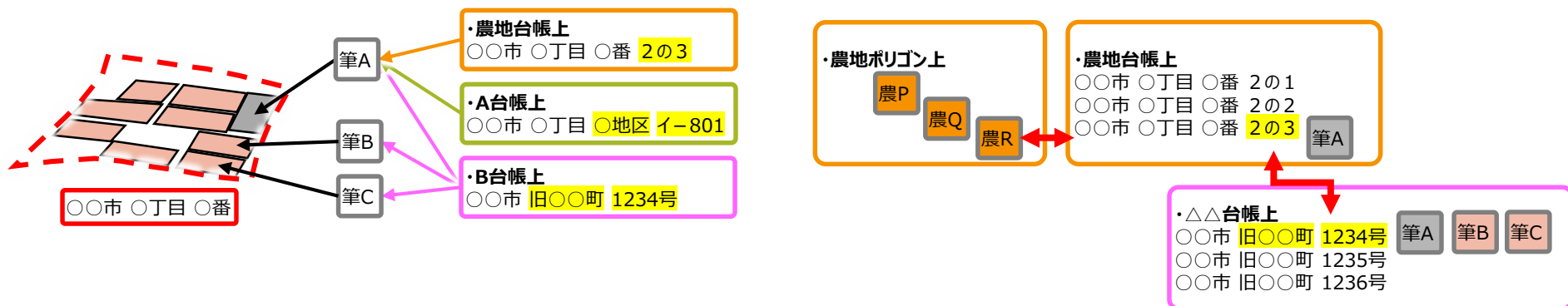
「デジタル地図」を活用した農地情報における期待・課題（2）

3. 農地台帳と筆ポリゴンとの紐付け

- IDや土地の履歴管理をどのように扱いやすく有するか
(データとしての持ち方の一般的な課題として)
- 同じ筆を表す複数の台帳データ（住所）のデータベース・デザイン

4. 最終的には農地台帳等の各台帳同士の紐付けまで

- ジオコーディングでは表記ゆれによる mismatching 等もあるので、台帳ごとの住所表記の細かさや難易度を元にユースケースを整理



おわりに

- 農地情報のデジタルデータは、地理空間情報*の活用全般において重要な資源
- ジオコーディング(逆ジオコーディング)の正確性を上げる上でも「辞書」にあたるデジタルデータとして台帳上の住所も同様に重要。
- 農地台帳等の各台帳同士の紐付けについては、筆ポリゴンや農地ポリゴンなど空間データの充実と同時に**データの履歴管理・ID体系の整理も重要**
- 複雑な管理体系や状況もあるので、ネット環境も含め比較的データ連携のしやすい市町村や集落でPoC的に検討・作業量の検証が必要

*: 地理空間情報活用推進基本計画(2017年3月)でも、農地情報や土地分類・土地利用情報が、他のデータとともにG空間社会の実現を下支えする地理空間情報である、と規定

以下、参考資料

<http://www.csis.u-tokyo.ac.jp/>

CSISのミッション

1. 空間情報科学の創生、深化、普及

- 先端的研究を実施して拠点の形成
- 空間情報科学の創成、深化、普及

2. 研究用空間データ基盤の整備

- 空間データを整備し、利用できるシステムを開発、実装し、空間データを利用する多くの分野の研究を支援

3. 産官学共同研究の推進

- 産官学の共同研究をはぐくむ場を提供

概ね1年を通していつでも申請可能

共同研究 申請からデータ入手すべてオンラインで完結



- 3月より継続申請が可能になりました。また、共同研究報告から）。
- 研究成果公開、共有のプラット

CSISでは、様々な空間データ「盤」として整備し、全国の研究者いただくことで、空間情報科学に

ただし、「研究用空間データ基
制限なく貸し出すことができない
業・団体から貸与されたデータも
同研究規定 (PDF) にのっとり、

- ▶ 号レベルアドレスマッチングサービス / Detailed Geocoding Service
- ▶ CSIS統計データベースサービス / CSIS Statistics Database Service
- ▶ 人の流れデータシリーズ / People Flow Project Series
- ▶ Zmap TOWN IIシリーズ / Zmap TOWN II Series
- ▶ 国勢調査シリーズ / Population Census of Japan Series
- ▶ 事業所・企業統計シリーズ / Establishment and Enterprise Census Series
- ▶ 経済センサスシリーズ / Economic Census Series
- ▶ 統計情報シリーズ / Statistics Information Series
- ▶ 国勢調査地図データシリーズ / GIS Data Series for Population Census of Japan
- ▶ アメダスシリーズ / AMeDAS Climate Series
- ▶ 気象データシリーズ / Climate Information Series
- ▶ 天気図シリーズ / Weather Chart Series
- ▶ GISMAPシリーズ / GISMAP Series
- ▶ 航空レーザ測量データシリーズ / Airborne Laser Scanning Series
- ▶ テレポイントシリーズ / Telepoint Phone/Address Series
- ▶ HD地形データシリーズ / High-Definition Topography Series
- ▶ マイクロジオデータシリーズ / Micro Geo Data Series
- ▶ 歴史GISデータシリーズ / Historical GIS Series
- ▶ 自然環境映像音声データシリーズ / Natural Environmental Video Data
- ▶ ダイナミックマップデータシリーズ / Dynamic map data series
- ▶ その他のデータセット / Others