

文章を自動的に地図化してリスクに対応する

オープンな地名情報処理システム GeoNLP の開発とリスク対応への活用

どんな研究？

- 文章中の地名を自動的に抽出
- 複数の候補があれば順位づけ
- 地図上に地名の分布を可視化
- 地名に ID を与えて統合

何がわかる？

- いまどこがどうなっている？
- 災害時などに迅速な状況把握
- アーカイブから地域情報へ
- どんなツイートが正しいか？

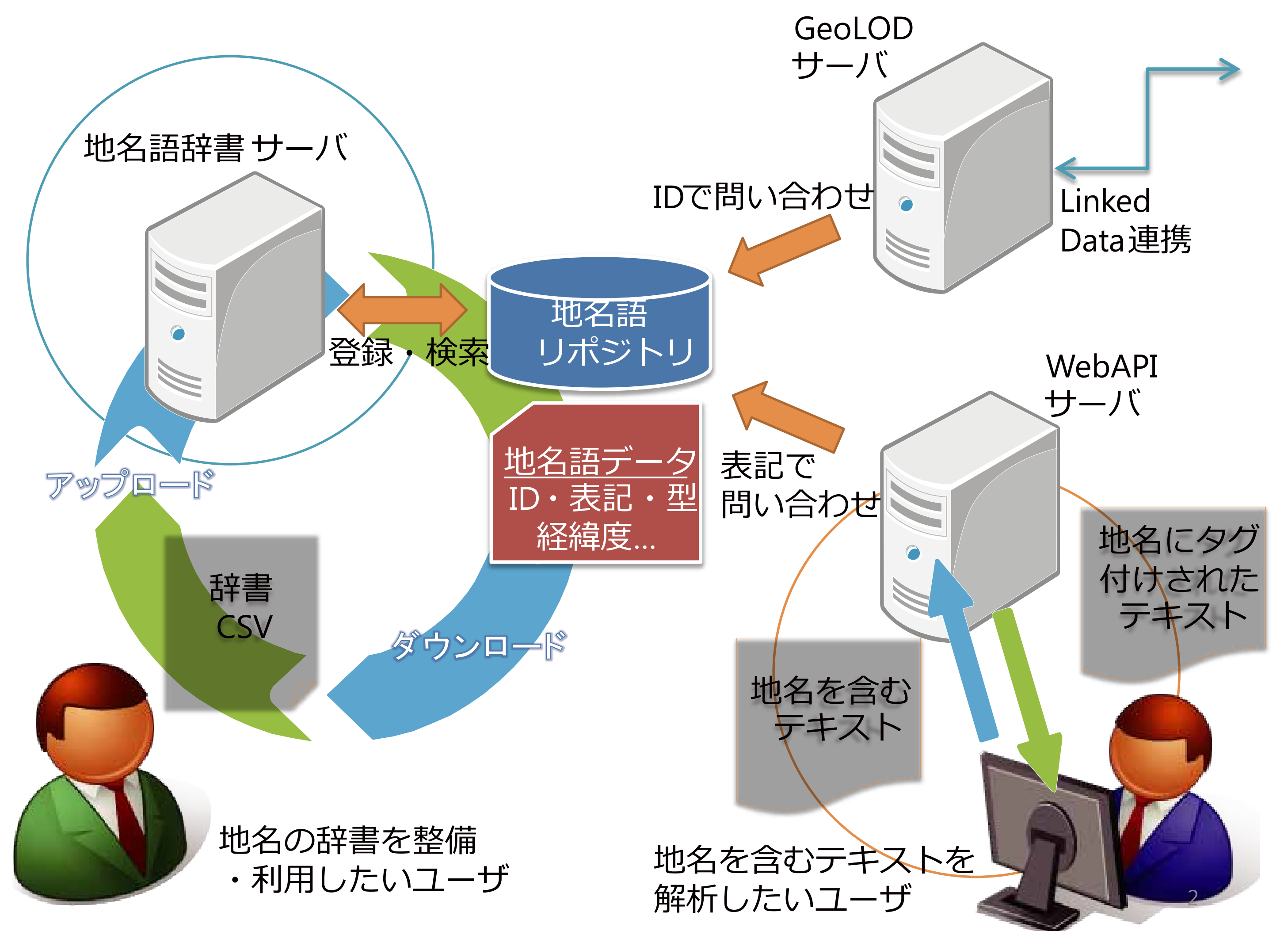
GeoNLP - オープンな地名情報処理システム <https://geonlp.ex.nii.ac.jp/>

地名を抽出・解決する

横浜は雨だよ。川崎は雪？

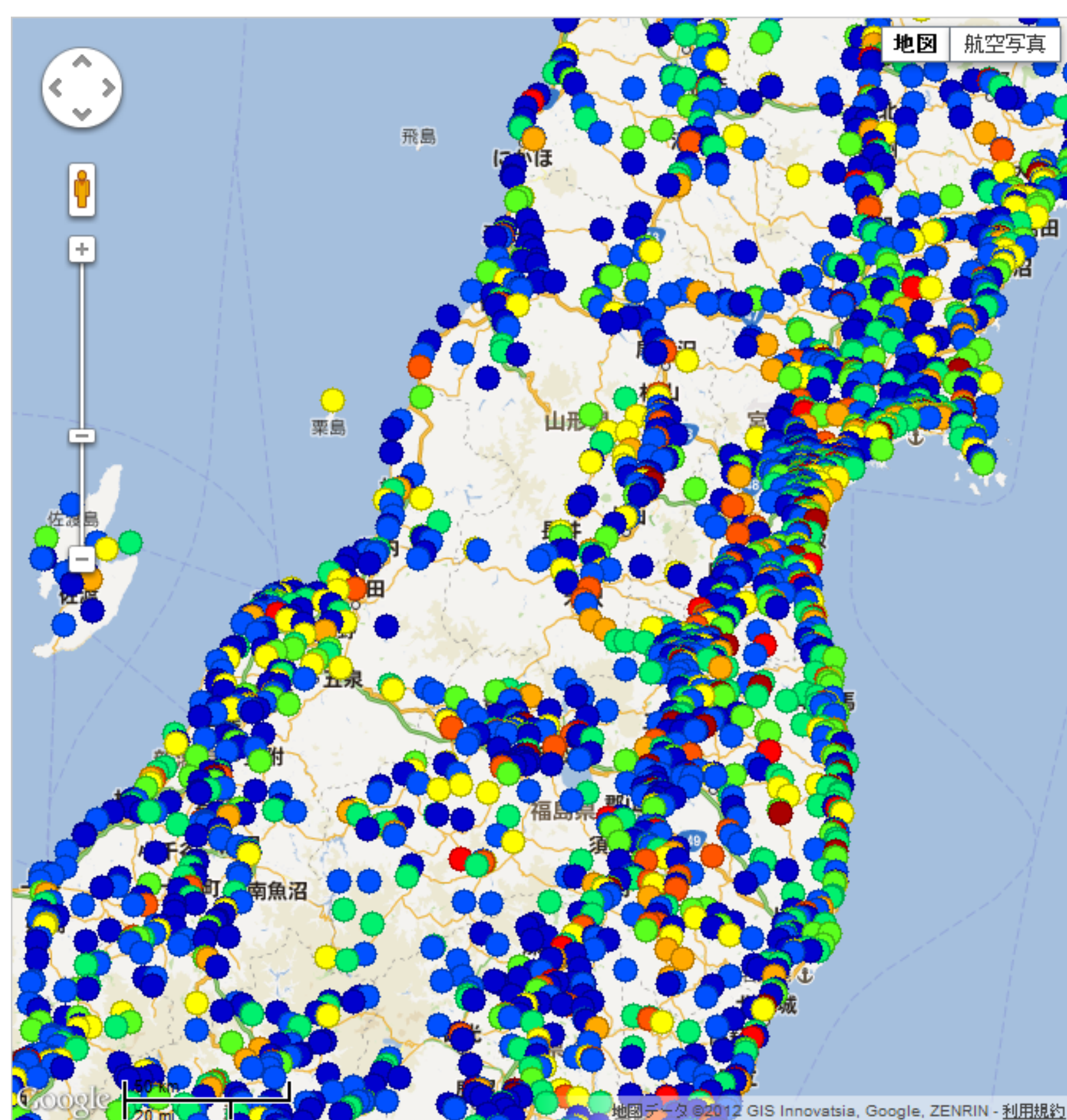
横浜は好きだよ。川崎は好き？

神奈川県横浜市と神奈川県川崎市？
 青森県横浜町と福岡県川崎町？
 横浜さんと川崎さん？
 地名の曖昧性、人名との判別などが問題。



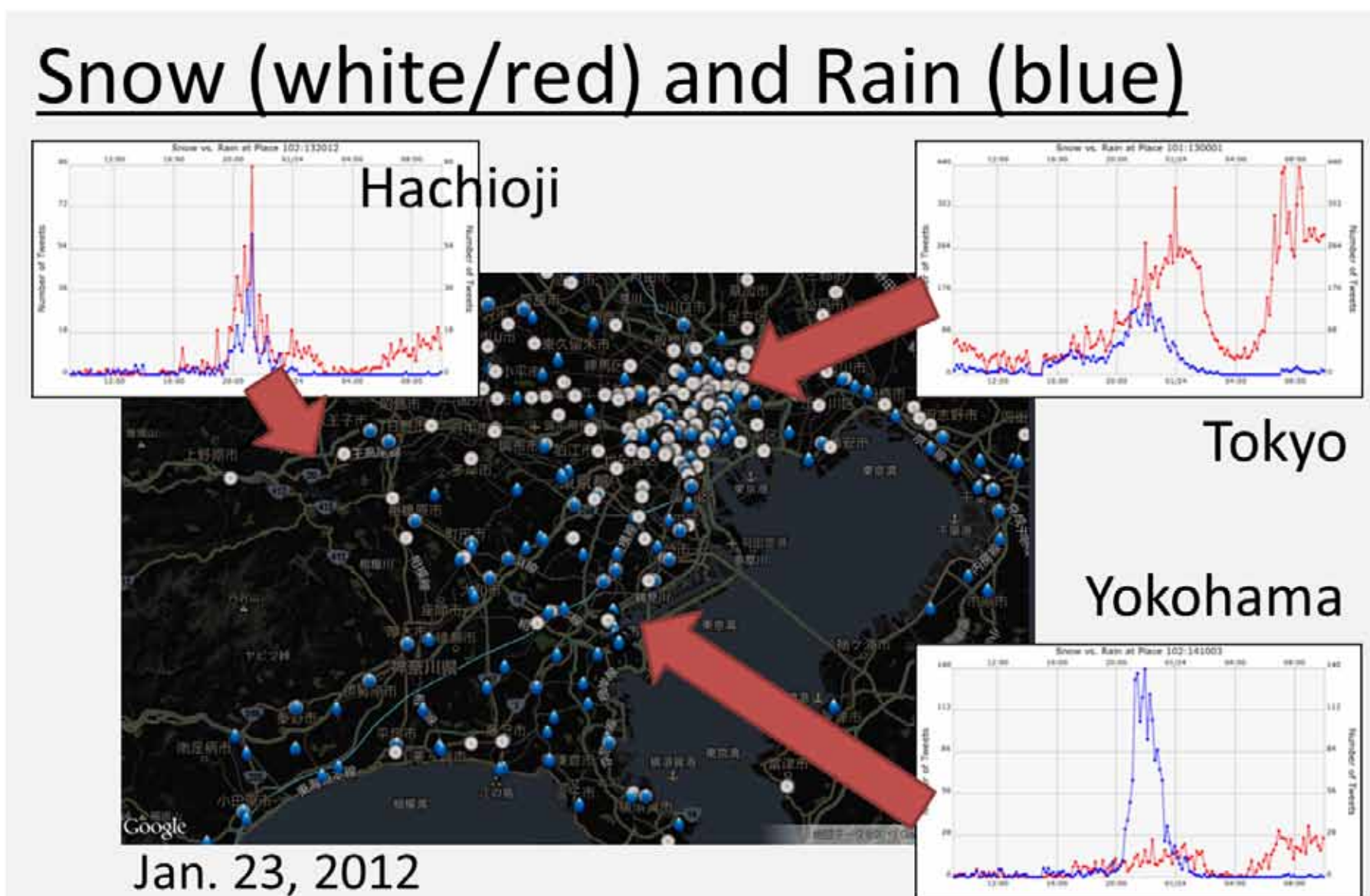
GeoNLP のリスク対応への応用と今後の課題

メディア解析



クライシス時の迅速な状況認識⇒自動テキスト仕分け

多メディアの比較



1. 「雨」「雪」「雹」などのキーワードを含むツイートを定期的に検索。
2. GeoNLP を使って地名を抽出。
3. ツイートの種類ごとにマッピングして降水状況を可視化。
4. レーダー等気象観測データと比較。

今後の課題

Ambiguity (曖昧さ)

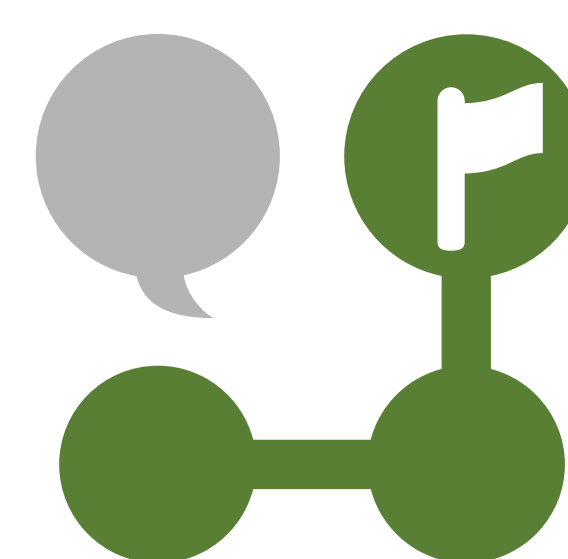
地名と非地名の区別に改良の余地が大きい。言語処理、機械学習、最適化等の手法活用？

Vagueness (曖昧さ)

「東京」はどの範囲を指すか？概念の定義(広がり)を学習し、スケールを含めて解釈。

オープンデータに基づく地名辞書

目的ごとに地名辞書の選択を可能とする、オープンデータの作成支援と流通促進。



総括：北本 朝展 (国立情報学研究所)
 GeoNLP Software 開発：相良 毅 (株式会社情報試作室)
 ウェブシステム構築：トライアックス株式会社
 協力：東京大学空間情報科学研究センター (CSIS)
 支援：地球環境情報統合プログラム (DIAS)